Chapter 34
Decision Theory: A Formal Philosophical Introduction

Richard Bradley

Abstract Decision theory is the study of how choices are and should be made.in a variety of different contexts. Here we look at the topic from a formal-philosophical point of view with a focus on normative and conceptual issues. After considering the question of how decision problems should be framed, we look at the both the standard theories of chance under conditions of certainty, risk and uncertainty and some of the current debates about how uncertainty should be measured and how agents should respond to it.

34.1 Introduction: Making Decisions

Decision problems abound. Consumers have to decide what products to buy, doctors what treatments to prescribe, hiring committees what candidates to appoint, juries whether to convict or acquit a defendant, aid organisations what projects to fund, and legislatures what laws to make. Descriptive decision theory aims to provide explanations for, and predictions of, the choices that are actually made by individuals and groups facing choices such as these. Normative decision theory, on the other hand, addresses the question of what decisions they should make and how they should make them: how they should evaluate the alternatives before them, what criteria they should employ, and what procedures they should follow. Our focus will be on the latter.

Decision problems arise for agents – entities with the resources to coherently represent, evaluate and change their environments in various possible ways – typically within the context of ongoing personal and institutional projects, activities or responsibilities. These projects together with the environment, both natural and social, provide the givens for the decision problems the agent faces: her resources for acting, her information and often her standards for evaluating outcomes, as well

R. Bradley (✉)
London School of Economics and Political Science, London, UK
e-mail: r.bradley@lse.ac.uk

S. O. Hansson, V. F. Hendricks (eds.), *Introduction to Formal Philosophy*, Springer Undergraduate Texts in Philosophy, https://doi.org/10.1007/978-3-319-77434-3_34

as the source of the problems she must respond to. Lastly, for agents to face a genuine decision problem they must have options: actions that they are capable of performing and equally of foregoing if they so choose. Some examples will illustrate the variety of forms such problems can take.

1. *Take a bus?* You have an appointment that you don't want to miss. If you walk you will arrive a little late. If you take the bus and the traffic is light, you should arrive a little ahead of time. On the other hand if the traffic is heavy then you will arrive very late, perhaps so late that the appointment will be lost. Is it worth risking it?
2. *Another slice of cake*. I have a weakness for chocolate cake which is contributing to a weight problem. My host offers me another slice of cake. Should I accept? I don't want to end up with diabetes or some other obesity related health problem, but one slice of cake will make very little difference and bring pleasure to both me and my host.
3. *Free condoms*. By supplying condoms free, rates of transmission of venereal disease can be considerably reduced. But there is the possibility that it will also encourage sexual activity thereby partially or even completely offsetting the benefits of a decreased transmission rate by virtue of the increase in the number of sexual liaisons.
4. *Road Building*. A new motorway linking two cities will reduce travelling time between the two of them and increase trade, with benefits for inhabitants of both cities. But those living close to the road will suffer from increased pollution and noise, as well as a fall in the value of their houses. Should it be built?

Many decision problems of the kind displayed in these examples can be described in the following way. A decision maker or decision making body has a number of options before them: the actions they can take or policies they can adopt. The exercise of each option is associated with a number of possible consequences, some of which are desirable from the perspective of the decision maker's goals, others are not. Which consequences will result from the exercise of an option depends on the prevailing features of the environment: whether traffic is light or heavy, how overweight I am, whether land prices are falling, and so on.

Let us call the set of environmental features relevant to the determination of the consequence of the exercise of any of the options, a state of the world. Then a decision problem can be represented by a matrix showing, for each available option, the consequence that follows from its exercise in each relevant state of the world. In our first example, for instance, taking the bus has the consequence of having to buy a ticket and arriving late in the event of heavy traffic and paying for a ticket and arriving early in the event of light traffic. This decision problem can be represented by a simple table such as the following:

More generally if A^1 through A^m are the m options open to the decision maker, s_1 through s_n are n possible states of the world (these must be mutually exclusive and exhaust all the possibilities), and C_1^1 through C_n^m are the $m \times n$ consequences that might follow from the choice, then a decision problem can be represented by a state-consequence matrix of the following kind:

	Heavy traffic	*Light traffic*
Take a bus	Arrive late Pay for a ticket	Arrive early Pay for a ticket
Walk	Arrive a little late No ticket needed	Arrive a little late No ticket needed

What choices should be made when facing a decision problem of this kind will depend on the circumstances the agent faces and in particular the amount of information she holds about its various features. Standard presentations distinguish between conditions of *certainty*, when the true state of the world, and hence the outcome of the action, is known; *risk* or *uncertainty*, when either the probabilities of the outcomes associated with an option are known (risk) or the agent can reach a judgement as to how probable they are on the basis of the information she holds (uncertainty); and *ignorance*, when nothing is known about the states. There are however many ways of butchering the beast and expositors draw the line between these conditions in different places (and indeed sometimes use these terms differently). Intermediate cases are important too, most notably when the decision maker is partially ignorant of the relevant probabilities – a situation commonly termed ambiguity.

When the decision maker knows the true state of the world, decision theory says that she should pick the option she considers best. When she is uncertain as to the actual state of the world, she must make a judgement as to how probable it is that each of the possible states is actually the case and pick the option whose expected benefit is greatest relative to these probability judgements. For instance suppose that I consider the probability of heavy traffic to be one-half and the benefit or desirability of the various possible consequences to be as below:

	Heavy traffic	*Light traffic*
Take a bus	-2	1
Walk	-1	-1

Then the expected benefit of taking the bus is a probability weighted average of the benefits of its possible consequences, i.e. $(-2 \times 0.5) + (1 \times 0.5) = -0.5$. On the other hand, walking has a certain benefit of -1. So in this case I should take the bus. But had the probability of heavy traffic been a lot greater, then walking would have been the better action.

More formally, let P be a probability measure on the states of the world and u a utility measure on consequences (we will say more about what these measures are and where they come from in due course). Then a state-consequence matrix,

Table 34.1
State-consequence matrix

	States of the world			
Options	s_1	s_2	...	s_n
A^1	C_1^1	C_2^1	...	C_n^1
...	...	...	...	...
A^m	C_1^m	C_2^m	...	C_n^m

such as that of Table 34.1, induces a probability-utility matrix in which options are represented as random variables that assign a utility value to each state of the world (intuitively, the utility of the consequence of exercising the option in question in that state).

So represented, each option has an expected value that is jointly determined by the functions u and P. The expected value of option A_1, denoted by $E(A_1)$ is, for instance, $u(C_{11}) \cdot P(s_1) + \ldots + u(C_{1n}) \cdot P(s_n)$. More generally, if the number of possible states of the world is finite[1]:

$$E(A^i) = \sum_{j=1}^{n} u(C_j^i) \cdot P(s_j)$$

Now what standard decision theory recommends is choosing the option with the highest expected value. This is known as the maximisation of expected utility hypothesis.

We will examine the maximisation hypothesis in greater detail later on. First, however, we look at a number of issues regarding the formulation and representation of decision problems. In the subsequent sections we look at the relation between preference and choice on the one hand and preference and utility on the other, setting aside complications arising from uncertainty. In the third section we return to decision making under uncertainty. In the final section we look at decision making under ignorance.

34.2 Framing Decision Problems

Decision theory makes a claim about what option(s) it is rational to choose, when the decision problem faced by the agent can be represented by a state-consequence matrix of the kind exemplified by Table 34.1. It is very important to stress that the theory does not say that you *must* frame decision problems in this way. Nor does it say that agents *will* always do so. It just says that *if* they are framed in this way, then only options which maximise expected benefit should be chosen. Nothing precludes the possibility that the same decision situation can or must be framed in different ways. This is true in more than one sense.

[1] The restriction to a finite number of states of the world is made for simplicity, but the expected value will still be well defined even if we drop it.

Firstly, it may be that the problem is not naturally represented by a state-consequence matrix. When I consider whether or not to have another slice of cake, for instance, it is not so much my uncertainty about the consequences of doing so that makes the choice difficult for me, but the contrast between desirability of the short term consequences (good) and the long-term ones (bad). So this problem should be given a different representation. We discuss this issue below in Sect. 34.2.

Secondly, the problem may not be representable by any kind of decision matrix at all because we are unable to identify the various elements of it: what our options are, what the relevant factors are that determine the outcome of each option, or what the consequences are of exercising one or another of the identified options when these factors are present. We discuss this problem in Sect. 34.4.

Thirdly, sometimes no structuring at all may be required; for instance, when certain actions are morally or legally obligatory or when habit dictates the course you take. These cases don't disprove the principle of maximising expected benefit. The point is rather that when the outcome of an action is certain, deliberation is redundant: the high probability of particular events or the great desirability (or otherwise) of particular consequences swamp the contribution that other factors might make.

34.2.1 Locations of Benefit

A two-dimensional decision matrix gives a two factor representation of a choice problem; in Table 34.1, for instance, these are just the states of the world and the consequences that follow from exercising an option in that world. But the state of the world in which a consequence is realised is not the only factor that matters to our assessment of its significance: this can also depend on who is affected by the actions and at what time and place. As John Broome [10] puts it, the good associated with an outcome of the exercise of an option has a number of different *locations*: people, places, times, qualities and states of the world. The desirability of being served cold beer, for instance, depends on the location of this service: it's good if the beer is served to me, in the evening, with a smile and when I have not had a few too many already; bad when it's for my children, or first thing in the morning, or during a philosophy lecture.

Locations of benefit are easily confused with *perspectives* on benefit, because many of the sorts of things that serve as the former, also serve as the latter. A perspective is a standpoint from which a judgement is made. You and I may reach different judgements because our standpoint differs: we might have different evidence and reasoning skills, perhaps different interests and biases, that lead us see things differently. Our standpoint also varies with time – as we get older, for instance, our aesthetic standards 'mature' – and sometimes with place and social role. But the way in which benefit varies with perspective need not be the same as the way it varies with location. I might now judge that it would be good if I were seen by a dentist next week. On the other hand, next week I might judge that the dentist is best avoided. Thus what I judge now to be the benefit next week of making an appointment now to see the dentist will be judged next week as anything but a benefit (even though the benefit, as judged from any temporal standpoint, does not depend when it obtains).

A consequence, in this more refined picture, is something that happens at a multi-dimensional location. Any one of these dimensions may be used to construct a two-factor matrix representation of a decision problem. For instance, when the problem is like the cake-eating one we can work with a time-consequence decision matrix like the following, in which the consequences of the relevant options (having or foregoing another slice of cake) at each relevant point in time are displayed.

	Times	
Actions	Now	Future
Another slice of cake	Pleasure from eating Host will be pleased	Risk of obesity and ill-health
Forego more cake	Forego pleasure Disappoint host	Likelihood of good health

A table of this kind makes it easy for the decision maker to focus on the question of how to weigh up present and future costs and benefits. Similar tables can be drawn up to assist reasoning with other locations, having different columns for the different people affected by the actions or the places at which the consequences occur, for instance. In the road building example for instance the salient locations are the people affected by the possible policy decisions. A person-consequence table helps the decision maker focus on the issue of the distribution of the benefits and costs to different people associated with each policy.

How decisions depend on the distribution of benefit across different dimensions of locations has been studied in different branches of decision theory: across states in the theory of decision making under uncertainty, across people in social choice theory, across time in intertemporal decision theory, across different qualities in multicriteria decision theory and so on. Moreover, the formal similarities between decision problems involving different locations has been a rich source of inspiration for decision theorists and has encouraged abstract examination of assumptions about the relationship between evaluations of consequences, locations and options. For the rest of this essay however I will focus on the decision problems in which uncertainty about the state of the world is the central feature. In fact the focus will be even more narrow than this, for I will say nothing about the very important case in which the events of which we are uncertain are the actions of other agents, leaving treatment of this to the chapter on game theory. Nonetheless many of the basic lessons drawn from the discussion here will apply in these other fields as well.

34.2.2 Choosing a Frame

It is typically possible to represent the decision problem one faces in more than one way: for instance, by location-consequence matrices that differ with respect to the locations they pick out or with regard to how finely the locations and consequences are individuated. In particular, they may be specified more or less finely or precisely, with the implication that a decision problem can always be refined (or coarsened) by adding detail to (or removing detail from) the description of the states and the consequence. This raises the question as to whether there are better and worse ways of representing a decision problem and if so, what these are. There are two claims that I want to make in this regard: firstly that not all representations of a decision problem are equally good and, secondly, that many representations are nonetheless permissible. This latter point is of some importance because it follows that an adequate decision theory must be 'tolerant' to some degree of the manner in which the problem is represented and that the solution it gives to a decision problem should be independent of how the problem is represented.

Let us start with the first claim, that some representations of a problem are better than others. A representation of a decision problem should help us arrive at a decision by highlighting certain features of the problem and in particular those upon which the decision depends. There are at least two considerations that need to be traded off when talking about the usefulness of a representation: the quality of the decisions likely to be obtained and the efficiency of obtaining them. To make a good decision, a decision maker must give appropriate weight to the factors upon which the decision depends. In deciding whether to take an umbrella or not, for instance, I need to identify both the features of the possible outcomes of doing so that matter to me (e.g. getting wet versus staying dry) and the features of the environment upon which these outcomes depend (e.g. the eventuality of rain). Furthermore I need to determine how significant these features are: how desirable staying dry is relative to getting wet, how probable it is that it will rain, and so on. If my representation of the decision problem is too sparse I risk omitting features that are relevant to the quality of the decision. If I omit possible weather states from my representation of the umbrella-taking decision, then I will fail to take into account factors – in particular the probability of rain – upon which the correctness of the decision depends. So, *ceteris paribus*, a representation that includes more relevant features will be better than one that does not.

One way of ensuring that no relevant features are omitted is simply to list *all* the features of possible outcomes and states of the world. But drawing up and making use of such a list is clearly beyond our human capabilities and those of any real agents. Reaching judgements costs in terms of time and effort. Representations that include too many features will result in inefficient decision making requiring more resources than is justified.[2] So, *ceteris paribus*, a simpler representation will be better than a more complicated one.

Achieving a good trade-off between accuracy and efficiency is not just a matter of getting the level of complexity right. It is also a matter of identifying the most useful features to represent explicitly. It is useful to represent a feature if it is (sufficiently) relevant to the decision and if we can determine what significance to attach to it. A feature of the state of the world or of a consequence is relevant to a decision problem if the choice of action is sensitive to values that we might reasonably assign to these features. For instance, whether it is desirable to take an umbrella with me or not will be sensitive to the probability of rain, but not sensitive at all to the probability of a dust storm on Mars.

The second aspect of usefulness is equally important. A representation should be appropriate to our informational resources and our cognitive capabilities in specifying features of the environment that we are capable of tracking and features of consequences that we are capable of evaluating. If the weather is relevant to my decision as to whether to take an umbrella or not, but I am incapable of reaching a judgement on the likelihood of rain or (perhaps I have no information relevant to the question or I don't understand the information I have been given) then there is little point in framing the decision problem in terms of weather contingencies. A good representation of a problem helps us to bring the judgements we are able to make to bear on the decision problem.

It follows of course that whether a framing is a useful one or not will depend on properties of the decision maker (and in more than one way). Firstly whether the features of the problem it represents are relevant depends on what matters to the decision maker and hence what sort of considerations her decisions will be sensitive to. And secondly whether a representation facilitates decision making will depend on the cognitive abilities and resources of the decision maker. Both of these will vary from decision maker to decision maker and from one time and context to another. It is clearly desirable therefore that a decision theory be 'representation tolerant' to as great a degree as possible, in the sense of being applicable to a decision problem irrespective of how it turns out to be useful for the decision maker to represent it.

34.3 Modelling Uncertainty

The modern theory of decision making under uncertainty has its roots in eighteenth century debates over the value of gambles, with Daniel Bernouilli (in [4]) giving the earliest precise statement of something akin to the principle of maximising expected utility. The first axiomatic derivation of an expected utility representation of preferences is due to Frank Ramsey [27] whose treatment in many way surpasses those of later authors. But modern decision theory descends from Savage, not

[2] What level of resources is justified will of course depend on what is at stake.

Ramsey, and it is in his book '*The Foundations of Statistics*' that we find the first rigorous simultaneous derivation of subjective probabilities and utilities from what are clearly rationality conditions on preference.

It is to Savage too that we owe the representation of the decision problem faced by agents under conditions of uncertainty that is now standard in decision theory. Savage distinguishes three types of object: states, consequences and actions. States of the world completely capture all the possible facts that might prevail in the decision situation that affect the outcome of acting. Consequences, on the other hand, are the features of the world that matter to the decision maker, such as that he is in good health or wins first prize in a beauty contest or is allowed to sleep late on a Sunday morning. Actions are the link between the two, the means by which different consequences are brought about in different states of the world. Formally, for Savage, they are just functions from states to consequences.

Although this tripartite distinction is natural and useful, Savage imposes some quite stringent conditions on these objects and the relationships between our attitudes to them. Firstly, states are required to be causally independent of the action the agent performs, while consequences are causally dependent on both the action and the state of the world. Secondly, the desirability of each consequence is required by Savage to be independent of the state of the world in which they are realised and of our beliefs about them, and vice versa (Binmore [5] calls this Aesop's principle). Both these conditions must hold if the representation of a decision problem by the kind of state-consequence matrix given in Table 34.1 can be transformed into a probability-utility matrix of the kind given by Table 34.2. The first ensures that the same probabilities can be applied to the states in comparing acts and the second that the utilities attached to consequences are state-independent.

The upshot is that Savage's theory is not partition independent in the way that I argued was desirable. Decision makers must represent the problems they face in a way which respects the conditions of probabilistic independence of the states from the acts and desirabilistic independence of the consequences from both the states and the acts. It is not always natural for us to do so. For instance in our earlier example of a decision as to whether to walk or take a bus we considered consequences such as paying for a ticket. But the desirability of such consequences are not state-independent. In particular they depend on all the possible contingencies that might arise, such as a medical emergency or an unexpected credit card bill, that

Table 34.2
Utility-probability matrix

	States of the world		
Options	$P(s_1)$	...	$P(s_n)$
A^1	$u(C_1^1)$	...	$u(C_n^1)$
...	...	...	...
A^m	$u(C_1^m)$	...	$u(C_n^m)$

require me to spend money. If too many of them arise a ticket would simply be unaffordable, if not many do it may be a trivial expense.[3]

34.3.1 State Uncertainty

A second feature of the representation of decision problems by a probability-utility matrices requires discussion. For Savage, an agent's uncertainty about what to do derives entirely from her uncertainty about what the state of the world is. This 'fundamental' uncertainty is captured by a probability function on the states of the world, measuring the degrees to which the agent judges or believes each state to be the actual one. There are two criticisms of this view of uncertainty that should be considered.

Firstly, the Savage model seems to ignore other forms of uncertainty and in particular the uncertainty that we might have regarding what value to attach to consequences and the uncertainty we might have regarding what actions are available. Both will be examined in more detail below

Secondly, there seems to be a significant difference between being unsure about when someone will arrive because one lacks precise information about their time of departure, traffic conditions, the route they have taken, and so on, and having absolutely no idea when they will arrive because you don't know when or whether they have left, whether they are walking or driving or indeed whether they even intend to come. In the former case, the information one holds is such as to make it possible to assign reasonable probabilities to the person arriving within various time intervals. In the latter, one has no basis at all for assigning probabilities, a situation of radical uncertainty that we previously termed *ignorance*. It may be rare for us to be totally ignorant, but situations of partial ignorance (or ambiguity), in which the decision maker is unable to assign determinate probabilities to all relevant contingencies, are both common and important.

More generally, according to some critics Savage's representation fails to distinguish between the different levels of confidence we might have, or have reason to have, in our probability judgements. Compare a situation in which we are presented with a coin about which we know nothing and one in which we are allowed to conduct lengthy trials with it. In both situations we might ascribe probability one-half to it landing heads on the next toss: in the first case for reasons of symmetry, in the second because the frequency of heads in the trials was roughly 50%. It seems reasonable however to say that our probability ascriptions are more reliable in the second case than the first and hence that we should feel more confident

[3] Savage was perfectly aware of this objection and drew an important distinction between small-world and grand-world decision problems. But he never produced a theory which, to his own and others satisfaction, explained how to convert grand-world problems into small-world ones satisfying the two requirements.

in them. To take this into account our state of uncertainty might be represented not by a probability function but by a set of reliability judgements over possible probabilities, or more formally, by a function $R : \Pi \to [0, 1]$, defined on a set $\Pi = \{p_i\}$ of probability functions on the set of events, and such that $\sum_i R(p_i) = 1$. These reliabilities could be incorporated into decision making in various ways, but the most natural perhaps is to prescribe choice that maximises reliability weighted expected utility. It is not difficult to see that such a rule is formally equivalent to maximising expected utility relative to the probability function $\bar{p} = \sum_i p_i.R(p_i)$. This is not an objection to introducing second-order probabilities, but merely to point out that use of reliabilities is more natural in the context of belief formation, than in decision making.

34.3.2 Evaluative Uncertainty

The distinctions between certainty, risk and uncertainty are standardly used only to characterise the agent's state of knowledge of the world. But it is equally important to distinguish cases in which consequences have known, or given, objective values and those in which these values are either unknown and the decision maker must rely on subjective evaluations of them, or do not exist and the decision maker must construct them. The possibility of evaluative uncertainty is typically ignored by decision theorists, because of their (often unconscious) attachment to the view that what makes a consequence valuable or otherwise (to the agent) is just that she desires it to some degree, or that she prefers it to a greater or lesser extent to other consequences. If this view were correct, talk of evaluative uncertainty would be misleading as one is not normally uncertain about what one's own judgement on something is (just what it should be).

There are however at least two ways in which one can be uncertain about the value to attach to a particular consequence or, more generally, whether one consequence is preferable to another. Firstly one may be uncertain about the factual properties of the consequence in question. If possession of the latest Porsche model is the prize in a lottery one is considering entering, one may be unsure as to how fast it goes, how safe it is, how comfortable and so on. This is uncertainty of the ordinary kind and, if one wishes, it can be 'transferred' (subject to some qualifications discussed in the next section) from the consequence to the state of the world by making the description of the consequence more detailed. For example, the outcome of the lottery may be regarded as having one of several possible consequences, each an instantiation of the schema 'Win a car with such and such speed, such and such safety features and of such and such comfort', with the actual consequence of winning depending on the uncertain state of the world.

Secondly one can be unsure as to the value of a consequence, not because of uncertainty about its factual properties, but because of uncertainty about whether these properties are valuable or as to how valuable they are. One may know all the specifications, technical or otherwise, of the latest Porsche and Ferrari models, so that they can be compared on every dimension, but be unsure whether speed matters more than safety or comfort. Once all factual uncertainty has been stripped from a consequence by detailed description of its features, one is left with pure value uncertainty of this kind.

When we assume that values are given, we take this uncertainty to have been resolved in some way. This could be because we assume that there is a fact of the matter as to how good a consequence is or as to whether one outcome is better than another, a fact that would be detailed by the true axiology. But it could also be because the description of the decision problem itself comes with values 'built-in'. For instance, in a problem involving a decision between two courses of medical treatment, it may be that a limited number of value considerations apply in the assessment of these treatments: number of patients saved, amount of discomfort caused, and so on. The decision theorist will be expected in such circumstances to apply only the relevant values to the assessment of the options, and to set aside any other considerations that he or she might 'subjectively' consider to be of importance. A large number of applications of expected utility theory take place in this sort of environment, when the issue of what values to apply have been settled by prior public policy debate.

In many situations, however, values are not given in any of these ways and the agent may be uncertain as to the value she should attach to the relevant prospects. In these circumstances the utility that the agent assigns to a consequence will reflect a subjective value judgement expressing her evaluative uncertainty. What kind of judgement this is a matter of considerable controversy, in particular regarding whether it expresses beliefs about factual properties of the consequences on which its desirability depends, beliefs about the objective normative properties instantiated by the consequences, or a judgement of a different kind to a belief. Formally, on the other hand, the only matter of concern is whether such judgements are adequately captured by utility ascriptions. If they are (as I believe), then considerations of evaluative uncertainty will have interpretative, but not formal, implications for expected utility theory. If not, new formal tools will need to be developed.

34.3.3 Option Uncertainty

In the state-consequence representation of a decision problem that we started with, actions were associated with definite consequences, one for each state of the world. But in real decision problems we are often unsure about the relationship between actions, worlds and consequences in essence because we do not know what consequence follows in each possible state of the world from a choice of action. For instance, we may be uncertain as to whether taking an umbrella will certainly have the consequence of keeping us dry in the event of rain. Perhaps the umbrella has holes, or the wind will blow it inside out or the rain will be blown in from the sides. We can put this difficulty in slightly different terms. If an action is *defined* as a particular mapping from states to consequences, then no uncertainty can arise about its consequences. But what we will then be unsure about is which actions are actually available to us i.e. which of the various hypothetical actions are real options. Whether we describe the problem as uncertainty about what options we have or as uncertainty about the consequences, in each state of the world, of exercising any of the options we know we have, is of little substance, and I shall use the same term – option uncertainty – to denote both.

Decision theorists tend to 'push' this uncertainty into the states of the world, by refining their description until all such contingencies are taken care of. They will regard a state of the world as insufficiently described by the absence or presence of rain, and argue that one needs to specify the wind speed and direction, the quality of the umbrella, etc. There are two reasons why this strategy will not work on all occasions. Firstly because, according to our best scientific theories, the world is not purely deterministic. When the conditions under which a coin is tossed do not determine whether a coin will land heads or tails, for instance, the act of tossing the coin does not have a predictable consequence in each state of the world. And secondly, even if we are in a purely deterministic set-up, it may be subjectively impossible for the decision maker to conceive of and then weigh up all the relevant contingencies or to provide descriptions of the states of the worlds that are sufficiently fine-grained as to ensure that a particular consequence is certain to follow, in each state, from the choice of any of the options open to them.

There are three strategies for handling this problem. One way is to use descriptions of the states of the world that identify the set of the conditions sufficient for the determination of the consequence, given the performance of the action, without actually enumerating the conditions. For instance, instead of defining actions in terms of states and consequences, we could take actions and consequences as our primitives and then define states of the world as consequence-valued functions ranging over actions. Similar strategies are advocated in the philosophical literature. Lewis [25], for instance, treats states as 'dependency hypotheses', which are just maximally specific propositions about how consequences depend causally on acts, while Stalnaker's [32] suggests that a state of the world be denoted by a conjunction of conditional sentences of the form 'If action A were performed then consequence C would follow; if action A' were performed then consequence C' would follow; if . . . '. By pursuit of any version of this strategy, option uncertainty is transformed into a particular kind of state uncertainty, namely uncertainty as to the true mapping from actions to consequences or as to the truth of the conjunction of conditionals that describes it.

A second strategy is to coarsen the description of the consequences to the degree necessary to ensure that we can be certain it will follow from the exercise of an option in a particular state. As Richard Jeffrey [18] points out, consequences may be identified by nothing more than act-state pairs, such as taking an umbrella in the rain and taking it in the snow. In his approach the outcomes of acts are taken to be *logical* consequences of act-state descriptions, but the coarsening of consequence-descriptions necessary to ensure option certainty need not be as radical as this.

Pursuit of this strategy converts option uncertainty, not into ordinary uncertainty about the state of the world, but into uncertainty about the desirability of the consequence as described – one part of what I previously called value uncertainty. We may be sure that the act of taking an umbrella will have the consequence in a rainy state of being able to protect ourselves against the rain by opening the umbrella. But whether this is a good thing or not depends on contingencies that by assumption we are unable to enumerate or identify. How bad it is to get soaked, for instance, depends on how cold the rainwater is. And rain temperature may be a variable about whose determinants we know very little. Whatever utility value we assign to the coarse-grained consequence of having an umbrella as rain-protection will embody this uncertainty and hence should be susceptible to revision.

The last strategy to consider, also originating in Richard Jeffrey's work, is the most radical and involves embracing option uncertainty rather than trying to reduce it to some other kind of uncertainty. This requires to think of actions not as functions from states to consequences, but as probability distributions over consequences. We will discuss this strategy in greater detail later on when presenting Jeffrey's theory.

34.4 Choice and Preference

Earlier we claimed that when a decision maker faces no uncertainty she should choose the option with the best consequences. There are two basic assumptions involved here. The first is Consequentialism: the idea that the only thing relevant to choice in these circumstances is the outcome or consequence of so choosing and not any feature of the means or process by which this outcome is achieved.[4] The second assumption is that there exists some value standard applicable to the outcomes which licenses talk of one or more of them being best. Jointly they entail that the decision maker ought to choose the action with the best consequence.

The value standard can have different interpretations, which in turn will imply different readings of the ought expressed by the choice principle. When the relevant standard is a subjective one, such as that based on the decision maker's preferences, the ought expresses a requirement of rationality, namely that she make a choice that is consistent with her subjective evaluation of its outcome. When the standard is an objective one, the prescription is to choose the action that has the outcome that is objectively best.

[4] It should be noted that the assumption of Consequentialism does not rule out a role for non-consequentialist considerations, in particular in determining the composition of the set of options. For instance if some actions are not permissable because they would violate someone's rights then they would be excluded from the option set. What it does assume is that such non-consequentialist considerations do not enter beyond this point.

34.4.1 Preference Relations

Let us try and make these claims more exact. First, some basic vocabulary. Let $X = \{\alpha, \beta, \ldots\}$ be a set of objects and let R be a binary relation on X. We say that R is:

1. *Transitive* iff for all $\alpha, \beta, \gamma \in X$, $\alpha R\beta$ and $\beta R\gamma$ implies that $\alpha R\gamma$ (and intransitive otherwise)
2. *Complete* iff for all $\alpha, \beta \in X$, $\alpha R\beta$ or $\beta R\alpha$ (and incomplete otherwise)
3. *Reflexive* iff for all $\alpha \in X$, $\alpha R\alpha$ (and irreflexive otherwise)
4. *Symmetric* iff for all $\alpha, \beta \in X$, $\alpha R\beta$ implies $\beta R\alpha$
5. *Antisymmetric* iff for all $\alpha, \beta \in X$, $\alpha R\beta$ and $\beta R\alpha$ implies that $\alpha = \beta$
6. *Acyclic* iff for all $\alpha_1, \alpha_2, \ldots, \alpha_n \in X$, $\alpha_1 R\alpha_2, \alpha_2 R\alpha_3, \ldots, \alpha_{n-1} R\alpha_n$ implies that not $\alpha_n R\alpha_1$.

In conditions of certainty, the assumption of Consequentialism implies that an option may be identified with the consequence of choosing to exercise it. So we can let the same set of alternatives represent both the options amongst which the agent must choose and the outcome of doing so. (A note of caution: to say that the consequence is certain is not to say that it is fully specified, so there may be disguised uncertainty.)

The decision maker's value standard is represented by a binary relation $\succeq$ on this set. Intuitively '$\alpha \succeq \beta$' means, on a subjective interpretation, that β is not preferred to α; on an objective one, that β is not better than α . In accordance with standard terminology I will call $\succeq$ a weak preference relation, without meaning thereby to impose a subjective interpretation. The strict preference relation $\succ$, indifference relation $\approx$, and comparability relation $\bowtie$, all on the set of alternatives X, are then defined by:

1. $\alpha \succ \beta$ iff $\alpha \succeq \beta$ and not $\beta \succeq \alpha$
2. $\alpha \approx \beta$ iff $\alpha \succeq \beta$ and $\beta \succeq \alpha$
3. $\alpha \bowtie \beta$ iff $\alpha \succeq \beta$ or $\beta \succeq \alpha$.

It will be assumed throughout that $\succeq$, $\approx$ and $\bowtie$ are all reflexive, that $\approx$ and $\bowtie$ are also symmetric, and that $\succ$ is a symmetric. It is common to assume that these relations are weak orders, i.e. that they are both transitive and complete. But the status of these two properties is very different. There are compelling grounds, on both subjective and objective interpretations, for assuming transitivity. Some authors have even argued that it belongs to the logic of comparative relations that they should respect it (e.g. Broome [10]). Completeness on the other hand cannot plausibly be said to be a requirement of rationality. Not only are we often unable to reach a judgement or don't need to, but on occasion it would be wrong of us to do so, e.g. when we expect to receive decisive information in the near future. Nor are there compelling grounds for supposing that objective betterness is complete: some goods may simply be incommensurable.

Why then do decision theorists so often assume completeness? One reason is that it makes the business of proving representation theorems for decision principles a lot easier mathematically speaking. A second reason lies in the influence of the Revealed Preference interpretation of decision theory. On this view having a preference for one alternative over another just is to be disposed to choose the former over the latter when both are available. Since agents are plausibly disposed one way or another in any choice situation (some choice is made after all), it follows that revealed preferences must be complete. But this interpretation has little to offer decision theory construed as either an explanatory or a normative theory. For if preferences are simply choice dispositions then citing someone's preferences cannot provide either an explanation or a justification of what they choose.[5]

The third argument, that completeness should be regarded as a requirement of coherent extendability, is the most cogent. The idea is this: although it is not a requirement of rationality that we should have reached a preference judgement regarding all prospects, it should nonetheless be possible to extend our current set of preferences to one which is both complete and consistent by reaching judgements about new prospects. If our current judgements are coherently extendible, then we can be sure that reaching new ones will not require a revision of our current preferences in order to retain consistency. Or to put it the other way round, if our preferences are not coherently extendible then as we reach judgements on prospects about which we formerly had no opinion, we run the risk of finding ourselves with an inconsistent set of preferences. Indeed we are sure to if we make enough new judgements. This does not give us a decisive reason to conform to the requirement of coherent extendability, as inconsistency can be avoided by revising some of our old judgements when we make new ones. But it does suggest that, *ceteris paribus*, it is pragmatically desirable to do so.

Suppose we accept the case for conformity with the requirement of coherent extendability. Then by studying the case of complete preferences we can derive a set of constraints on our beliefs and desires that must be fulfilled in order that they too be coherently extendible. For instance, if we can show that the rationality of a complete set of preferences implies that our beliefs must have some particular property P, then we can conclude that our (incomplete) beliefs must have the property of being extendible to a set of beliefs having P.

34.4.2 Choice

Let X be a finite set of alternatives and C be a choice function on $\wp(X)$: a mapping from subsets $A \subseteq X$ to subsets $\varnothing \subset C(A) \subseteq A$. The choice function C will be said to be *specific* iff its range is restricted to singleton sets. Intuitively $C(A)$

[5] To be clear, it is the ontological doctrine just described that should be rejected, not the associated epistemological doctrine according to which knowledge of preferences ultimately rests on observations of choice. The latter, in contrast to the former, has much to recommend it.

is the set of objects from the set A that could be chosen: could permissibly be so in normative interpretations, could factually be so in descriptive ones. When C is specific a further interpretation is possible, namely that $C(A)$ is the object observed to be chosen from the set A.

We are especially interested in the case when a choice function C can be said to be based on, or determined by, a weak preference relation $\succeq$. A natural condition for this being the case is that an object is chosen from a set if and only if no other object in the set is strictly preferred to it. Formally:

$$(\textbf{PBC})\ \alpha \in C(A) \Leftrightarrow \neg\exists\beta \in A : \beta \succ \alpha$$

PBC is sometimes called the *Maximality* condition. With a qualification that will be made a little later on, PBC seems necessary for preference based choice. But is it sufficient? Sen [30] suggests to the contrary that it is not enough that nothing be (comparably) better than what is chosen, it must also be the case that what is chosen is (comparably) no worse than any alternative. More formally, preference-based choice should satisfy Strong PBC or as it is more commonly called:

$$(\textbf{Optimality})\ \alpha \in C(A) \Leftrightarrow \forall\beta \in A, \alpha \succeq \beta$$

To examine these proposals let us use the weaker criterion of maximality to derive a set-valued function on $\wp(X)$ from the agent's preferences by defining, for all $A \in \wp(X)$:

$$C_{\succeq}(A) := \{\alpha \in A : \neg\exists\beta \in A,\ \beta \succ \alpha\}$$

Then:

Theorem 1

(a) $C_{\succeq}$ is a choice function iff $\succeq$ is acyclic
(b) Choice function $C_{\succeq}$ satisfies Optimality $\Leftrightarrow$ $\succeq$ is complete.
(c) Choice function $C_{\succeq}$ is specific $\Leftrightarrow$ $\succ$ is complete.

The proof of (a) and (b) can be found in Sen [30], (c) follows immediately.

Two comments. Firstly, Theorem 1(b) shows that to require satisfaction of Strong PBC is to make completeness of an agent's preferences a condition for their choices to be preference-based. But this seems unreasonable. As we have seen, completeness has little normative appeal as a preference condition and someone with incomplete preferences whose choices satisfy PBC can be said to be making these choices in the light of their preferences to the maximum extent possible. On the other hand, as Theorem 1(c) shows, neither satisfaction of PBC nor of Strong PBC is sufficient for preference to determine the choice of a specific alternative. For when two alternatives are incomparable or indifferent then both are permissible choices. The upshot is that we should regard satisfaction of PBC as the mark of preference-based choice, noting that only when an agent's *strict* preferences are complete will this condition suffice for preference to determine choice completely.

Secondly, there are various ways of giving substance to the notion of being preference-based. On an *explanatory* reading, it means that the decision maker's preferences explain the choices that she makes by providing the reasons for them. On the other hand, on a *normative* reading, it means that the decision maker's preferences rationalise or justify the choices that she makes. Revealed Preference theorists regard neither of these interpretations as warranted and advocate a third, purely *descriptive* reading, according to which 'preference-based' means no more than that a choice function can be represented by a preference relation. The first two interpretations give primacy to preferences, with PBC doing service as a principle by which we infer properties of choice from properties of preferences. The last interpretation, on the other hand, gives primacy to the properties of choices and to the problem of deriving properties of preferences from them.

The main condition of Revealed Preference theory is the Weak Axiom of Revealed Preference (WARP), which says that if α should be chosen from a set containing β, then whenever β should be chosen and α is available, α should also be chosen. Formally, we follow Sen [30] in breaking this down into two conditions:

Axiom 2

(WARP) Suppose $\alpha, \beta \in B \subseteq A$. Then:
(Condition Alpha) If $\alpha \in C(A)$, then $\alpha \in C(B)$
(Condition Beta) If $\alpha, \beta \in C(B)$ and $\beta \in C(A)$, then $\alpha \in C(A)$

Theorem 3 *Let C be a choice function on $\wp(X)$. Then:*

(a) C satisfies Alpha if there exists a relation $\succeq$ on X such that $C = C_{\succeq}$
(b) C satisfies WARP iff there exists a weak order $\succeq$ on X such that $C = C_{\succeq}$.

Proof (a) Suppose that there exists a relation $\succeq$ on X such that $C = C_{\succeq}$, that $\alpha \in B \subseteq A$ and that $\alpha \in C(A)$. By definition $\forall\beta \in A, \alpha \succeq \beta$ or $\beta \not\succeq \alpha$. Hence $\forall\beta \in B$, $\alpha \succeq \beta$ or $\beta \not\succeq \alpha$. Then by definition, $\alpha \in C(B)$. (Note that the converse is not true: it does not follow that if C satisfies Alpha that $C_{\succeq}$ is a choice function.) The proof of (b) can be found in Sen [30]. ■

Theorem 3(b) seems to give Revealed Preference theory what it needs, namely a characterisation of both the conditions under which the agent's preferences are 'revealed' by her choices and of the properties of these preferences. In particular if her choices respect WARP then a transitive and complete weak preference relation can be imputed to her which, together with PBC, determines these choices. But this observation is of very little normative significance in the absence of a reason for thinking that choices should satisfy the WARP axiom. The problem is that, unless $\succeq$ is complete, a preference-based choice function need not satisfy condition Beta. Suppose, for example, that the agent cannot compare α and β, but that no object in B is preferred to either. So both are permissible choices. Now suppose that $A = B \cup \{\gamma\}$ and that γ is preferred to α but not comparable with β. Then β is a permissible choice but not α. Since it is no requirement of rationality that preferences be complete, I take it that WARP is not normatively compelling. Hence preferences are not fully revealed by the choices they determine.

Condition Alpha is sometimes called the Independence of Irrelevant Alternatives condition in view of the fact that it implies that the framing of the choice set shouldn't influence an agent's preferences. The fact that it is implied by PBC, in the sense given by Theorem 3(a), is grounds for thinking it should be respected by choices. But as Sen has pointed out, the composition of the choice set itself can matter. When offered the choice between staying for another drink or leaving the party, I might choose to stay. But if offered the choice between leaving the party, staying for a drink or staying to participate in a satanic ritual I may well choose to leave.

It seems therefore that what preference-based choice requires is something more subtle than picking non-dominated alternatives relative to a given preference relation. It is this: that we should not choose any alternative from a set, when there is another in that set that it strictly preferred to it, *given* the set of alternatives on offer. Making this criterion for preference-based choice formal is tricky. Nonetheless, as we shall see later on, it has important conceptual implications.

34.5 Utility Representations

Preference relations that are weak orders can be represented numerically, thereby allowing for an alternative characterisation of rational choice. More exactly, let us call a function $U : X \to \mathbb{R}$, a *utility representation* of the weak order $\succeq$, iff for all $\alpha, \beta \in X$:

$$\alpha \succeq \beta \Leftrightarrow U(\alpha) \geq U(\beta)$$

Then:

Theorem 4 *Suppose that the preference relation $\succeq$ is a weak order on a countable set X. Then there exists a function U that is a utility representation of $\succeq$. Furthermore U' is another such a utility representation iff U' is a positive monotone transformation of U i.e. there exists a strictly increasing function $f : \mathbb{R} \to \mathbb{R}$ such that $U' = f \circ U$.*

See Kranz et al. [23, Section 2.1] for a proof of this theorem. In case X is not countable, numerical representability is not assured for weak orders unless X has a 'dense' subset – one containing elements lying (in the weak order) between any two prospects in X. When the preference order is lexicographic for instance this condition will be violated. In contrast, any continuous weak relation on a connected topological space *is* numerically representable by a continuous function (see Kreps [24] for details), where the continuity of a relation is defined as follows:

Continuity: For any subset $\{\alpha_i\}$ such that $\alpha_1 \models \alpha_2 \models \ldots \models \alpha_n$ and $\beta \succsim \alpha_n \succsim \gamma$, $\beta \succsim \alpha_i \succsim \gamma$, for all large i.

These representation results have an obvious weakness: the assumption that preferences are complete. But it is in fact simple enough to generalise the result to all transitive preference relations, complete or otherwise, by defining a utility representation of the transitive relation $\succeq$ to be a set $\mathcal{U}$ of utility functions such that for all $\alpha, \beta \in X$, $\alpha \succeq \beta$ iff for all $U \in \mathcal{U}$, $U(\alpha) \geq U(\beta)$. Such a set $\mathcal{U}$ may be constructed by placing in it, for each possible 'completion' of $\succeq$, a utility function that represents the resultant weak order. It follows that the set will inherit the uniqueness properties of its elements i.e. that $\mathcal{U}$ will be unique up to positive monotone transformation. More formally, let us say that a preference relation $\succeq$ on a set X is represented by a set of real-valued functions Φ just in case for all $\alpha, \beta \in X$,

$$\alpha \succsim \beta \Leftrightarrow F \in \forall F \in \Phi, F(\alpha) \geq F(\beta)$$

Then:

Theorem 5 (Evren and OK [14, p. 5]) *Let $\succsim$ be a weak order on a set X. Then there exists a set Φ of real-valued functions that represents $\succeq$.*

Theorem 4, together with the discussion in the previous section, implies that choices that are preference-based, in the sense of satisfying Optimality, are utility maximising. But one must be careful not to attach too much significance to this characterisation of utility maximisation. The mere existence of the function U that represents preferences does not in itself explain the agent's preferences, nor does it justify them. It merely describes them numerically. The contrast with belief is instructive. Under certain conditions (which will be described later on), a correspondence can be established between beliefs and preferences over specific alternatives, namely those whose desirability depends on the truth of the contents of the beliefs in question. In this case we are inclined to speak of the beliefs being the cause or reason for the preference. This is because we have independent scientific grounds for attributing causal powers to beliefs. Similarly for preferences. But we have no such grounds for attributing causal or justificatory powers to the utilities of alternatives distinct from the agent's preferences for them. We might speak, as I will do, of a utility judgement and the considerations upon which it is based. But this is no more than shorthand for talk of preferences, in which transitivity and completeness are taken for granted. Such talk has its dangers: in particular it can encourage one to read more into the numbers than is allowed by the representation. But it is also convenient; hence our interest in being clear about their content.

Theorem 4 is rather weak, as the uniqueness properties of the utility representation it establishes make manifest. With stronger assumptions about the set of alternatives and preferences over them, more interesting properties of the utility representation can be derived and its uniqueness increased. In the next couple of sections we will characterise the conditions on weak orders under which there exists an additive utility representation of it. Since our primary interest is in the normatively significant properties of preference relations and corresponding utility representations, and not with problem of numerical representation itself, I will be somewhat cavalier in my statement of the more technical conditions on preference orders. For example, to obtain a cardinal representation of a weak order it is typically necessary to assume an Archimedean condition; to ensure, in effect, that any two objects in the domain of the weak order are comparable, no matter how far apart they lie in that order. The exact condition required will depend on the nature of the set of alternatives being assumed, and for this reason I will not spell it out each time. For the details on these, the most comprehensive source is Krantz et al. [23].

34.5.1 Conjoint Additive Structures

For a first extension we return to our initial representation of a decision problem as a matrix of locations and consequences. The objects of choice here are ordered sets of outcomes, one for each possible state of the world or, more generally, location. Consequently the set of alternatives forms a product set of the form $X = X_1 \times X_2 \times \ldots \times X_n$, where each X_i is the set of possible outcomes at the ith location. A profile $(x_1, x_2, \ldots, x_n) \in X$ could be a set of attributes of a good, for instance, or a set of allocations to individuals, or a set of events at different times.

This structure allows for stronger assumptions about rational preference and a correspondingly richer utility representation of them. For any subset K of the set of possible locations, let $X_K := \prod_{j \in K} X_j$. For any partitions $\{K, L\}$ of the set of locations, let (a, c) be the member of X where $a \in X_K$ denotes the values of the locations in K and $c \in X_L$ denotes the values of the locations in L. Then consider:

Axiom 6 (Strong Separability) *For all partitions {K,L} of the set of locations and for all $a, b \in X_K$ and $c, d \in X_L$:*

$$(a, c) \succeq (b, c) \Leftrightarrow (a, d) \succeq (b, d)$$

The axiom of strong separability appears in different contexts under a wide variety of names, most notably Joint Independence [23] and the Sure-thing Principle [29] for the case where locations are states. It has a strong claim to be the most interesting and important of the conditions regularly invoked by decision theorists. On the one hand, it does not have the same normative scope as the transitivity condition. For instance, consider its application to allocations to different individuals. In this context Strong Separability rules out a direct sensitivity to inequality, such as might be manifested in a preference for (a, a) over (b, a) and for (b, b) over (a, b). Similarly in applications to decisions with outcomes at different temporal locations it rules out a preference for novelty over repetition, such as might be manifested in a preference for (a, b) over (b, b) and for (b, a) over (a, a). On the other hand, in many applications and when outcomes are carefully described, the axiom does seem normatively compelling. We return to this in the discussion of decision making under uncertainty.

Let us say that a location l is *essential* just in case there exist x_l and y_l such that $(x_1, .., x_l, \ldots, x_n) \succ (x_1, .., y_l, \ldots, x_n)$ for all $x_1, \ldots, x_n \in X_{N-l}$. And let us say that $\succeq$ is *solvable* on X just in case $(x_1, .., \bar{x}_l, \ldots, x_n) \succ (y_1, .., y_l, \ldots, y_n) \succ (x_1, .., \underline{x}_l, \ldots, x_n)$ implies that there exists an x_l such that $(x_1, .., x_l, \ldots, x_n) \approx (y_1, .., y_l, \ldots, y_n)$. If the preference relation $\succeq$ is Archimedean and solvable on X, then we call the pair $\langle X, \succeq \rangle$ an additive conjoint structure. Then:

Theorem 7 *Let $\langle X, \succeq \rangle$ be an additive conjoint structure with at least three essential locations. Assume that $\succeq$ satisfies Strong Separability. Then there exists a utility representation U of $\succeq$ on X such that for all $(x_1, x_2, \ldots, x_n) \in X$:*

$$U(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} u_j(x_j)$$

for some family of functions $u_j : X_j \to \mathbb{R}$. Furthermore if U' and the u'_j are another such a family of utility representations, then there exists constants $a, b, a_j, b_j \in \mathbb{R}$ such that $U' = aU + b$ and $u'_j = a_j.u_j + b_j$.

The proof of this theorem involves three main steps (see Krantz et al. [23] for details). First, we observe that by application of Theorem 4, there exists a utility representation U of $\succeq$ on X, unique up to positive monotone transformation. The second step is to derive location-relative preference relations from $\succeq$, in which essential use is made of Strong Separability. In the light of Theorem 4 this implies the existence of location-relative utility functions – the u_j – also unique up to positive monotone transformation. The final step is to show that judicious choice of scales for the u_j permits U to be expressed as a sum of them.[6]

Theorem 7 has many applications. For a historically important example suppose that the X_j are different individuals and the x_j allocations that are made to them. Then Theorem 7 asserts the existence of an additive utility representation of any set of strongly separable preferences over allocations to individuals. This is typically called a utilitarian representation of social decisions.

34.5.2 Linear Utility

We now consider an even richer structure on the objects and a stronger restriction on preferences sufficient to ensure the existence of a linear representation of them.

[6] It is important to note that it's essential to the possibility of an additive representation that no cross-locational comparisons are possible. For such comparisons would constrain the co-scaling of the u_j and there would then be no guarantee that the permitted co-scaling allowed for an additive representation.

A set of objects X is said to be a *mixture set* iff for all $\alpha, \beta \in X$ and any $0 \leq k \leq 1$, there exists an element in X, denoted by $k\alpha + (1-k)\beta$, such that:

1. If $k = 1$ then $k\alpha + (1-k)\beta = \alpha$
2. $k\alpha + (1-k)\beta = (1-k)\beta + k\alpha$
3. For all $0 \leq l \leq 1$, $l(k\alpha + (1-k)\beta) + (1-l)\beta = lk\alpha + (1-lk)\beta$

Axiom 8 (Linearity) *For all $\alpha, \beta, \gamma \in X$ and any $0 \leq k \leq 1$:*

$$\alpha \approx \beta \Leftrightarrow k\alpha + (1-k)\gamma \approx k\beta + (1-k)\gamma$$

Axiom 9 (Archimedean) *For all $\alpha, \beta, \gamma \in X$, if $\alpha \succ \gamma \succ \beta$ then there exist k and l such that:*

$$k\alpha + (1-k)\beta \succ \gamma \succ l\alpha + (1-l)\beta$$

Theorem 10 (Herstein and Milnor [17]) *Assume that X is a mixture set and that $\succeq$ is an Archimedean weak order on X that satisfies the Linearity axiom. Then there exists a utility representation U of $\succeq$ on X such that for all $\alpha, \beta \in X$:*

$$U(k\alpha + (1-k)\beta) = kU(\alpha) + (1-k)U(\beta)$$

Furthermore U' is another such a utility representation iff U' is a positive linear transformation of U i.e. there exists constants $a, b \in \mathbb{R}$ such that $U' = aU + b$.

One very important application of the idea of a mixture space is to lotteries. Let Z be a (finite) set of outcomes or ‘prizes’ and let the set of lotteries $\Pi = \{p_i\}$ be a set of a probability mass functions on these outcomes i.e. each $p_i \in \Pi$ is a function from the $z \in Z$ to the interval $[0, 1]$ such that $\sum_z p_i(z) = 1$. For any $p_i, p_j \in \Pi$, let $kp_i + (1-k)p_j$, called the k-compound of p_i and p_j, denote the member of Π defined by:

$$(kp_i + (1-k)p_j)(z) := kp_i(z) + (1-k)p_j(z)$$

It follows that Π is a mixture set of lotteries.

When applied to lotteries the Linearity axiom is typically called the Independence axiom: it says that if two lotteries p_i and p_j are equally preferred then a k-compound of p_i and p_k is equally preferred to a k-compound of p_j and p_k. The Archimedean condition amounts to saying that no matter how good p_i is (how bad p_j is) there is some compound lottery of p_i and p_j which gives p_i such small (large) weight that p_k is strictly preferred to it (it is strictly preferred to it p_k). Or more pithily, everything can be traded off if the probabilities are right.

Theorem 11 (Von Neumann and Morgenstern) *Let $\succeq$ be an Archimedean weak preference order on Π that satisfies the Independence (Linearity) axiom. Then there exists a utility representation U of $\succeq$ on Π and a function $u : Z \to \mathbb{R}$ such that for all $p_i \in \Pi$:*

$$U(p_i) = \sum_{z \in Z} p_i(z).u(z)$$

See Kreps [24] for an instructive proof of this result. Von Neumann and Morgenstern's theorem is usually considered to belong to the theory of decision making under uncertainty and its appearance here bears out my earlier claim that the distinction between certainty and risk is a matter of perspective. When making decisions under risk we *know* what situation we face, even if we don't know what the final outcome will be. This makes it a convenient bridgehead to the topic of uncertainty.

34.6 Decisions Under Uncertainty

It is now time to make more precise the claim that in situations of uncertainty, choices should maximise expected utility. Although this prescription is still consequentialist in spirit the explicit introduction of uncertainty requires a more nuanced expression of what Consequentialism entails in these circumstances. More specifically, in these circumstances, the choice-worthiness of an action depends not only on the consequences of the action but also on the relative likelihood of the possible states of the world in which the consequences might be realised. The prescription to maximise expected utility is made relative to a specification of the probabilities of states of the world and utilities of the consequences. There are thus two relations that need to be examined: the value relation that we discussed before and a possibility or probability relation on the states of the world expressing the decision maker's state of uncertainty. Both the properties of these relations and of the quantitative representations of them are relevant to the derivation of the expected utility principle.

Like the value ordering, the possibility ordering can be given both a subjective and objective interpretation, as can the numerical probabilities based on it. This means that in principle the prescription to maximise expected utility is amenable to four different readings with quite different normative implications. If both are construed objectively (as in, for instance, Broome [10]) then the principle prescribes action which maximises the objective expectation of goodness. If preferences are subjective but probabilities are objective (as they are in Von Neumann-Morgenstern decision theory [35]) then the principle prescribes maximisation of the objective expectation of subjective preference. If both are construed subjectively (as in Savage [29]) then the prescription is to maximise the subjective expectation of subjective preference and so on.

As the normative claims of these different interpretations of expected utility theory are rather different, one should not expect that one type of argument will serve to justify all of them. What we can do however is to build a common platform for such arguments by identifying the properties of the two ordering relations that are necessary and sufficient for the existence of an expected utility representation that justifies (either by rationalising or by normatively validating) the decision maker's choice. By an expected utility representation, I mean an assignment of utilities to consequences and probabilities to states of the world such that the agent's preferences over options cohere with their expected utility.

More formally, let Ω be a set of consequences, $S = \{s_{1,}s_{2,} \ldots\}$ be a set of states of the world and $\mathcal{F} = \{A, B, C, \ldots\}$ be the set of subsets of S, called events. Finally let $\Gamma = \{\alpha, \beta, \gamma, \ldots\}$ be the set of actions, where an act is function from S to Ω. In the light of an earlier remark that the difference between states and consequences is pragmatic rather than ontological, it makes sense to treat the latter as a type of event, rather than following Savage in treating them as logically distinct. Formally this means that $\Omega \subseteq \mathcal{F}$.

Let $\succeq$ be a preference relation on the set of actions. A function $V : \Gamma \to \mathbb{R}$ is called an *expected utility representation* of $\succeq$ iff V is a utility representation of $\succeq$ and there exists a real valued function $u : \Omega \to \mathbb{R}$ and a probability function $P : \mathcal{F} \to \mathbb{R}$ such that for all $\alpha \in \Gamma$:

$$V(\alpha) = \sum_{s_i \in S} P(\{s_i\}).u(\alpha(s_i))$$

Our examination will be conducted in two steps. In the first we apply the Von Neumann and Morgenstern theory to decision making under risk, i.e. to conditions in which probabilities are given. And in the second we present Savage's derivation of such a probability from the agent's preferences.

34.6.1 Expected Utility Theory

Suppose that our decision problem takes the form given by Table 34.1. We want to know under which conditions a preference relation over the available options has an expected utility representation. Consider first a situation in which the probabilities of the states of the world are known, a circumstance to which Von Neumann-Morgenstern utility theory is usually applied. It is important to note that to do so we must assume that the decision problem we face can be adequately represented as a choice between lotteries over outcomes. For this it is not enough that we know the probabilities of the states, we must also assume that the only feature of these states which matters for our choice of action is their probability. In particular, the fact that an outcome is realised in one state or another must not influence its desirability. This is known as the assumption of *state-independence*. It appears in an explicit form in the axiomatisations of expected utility theory given by Savage and by Anscombe and Aumann, but is merely implicit in the adoption of the Von Neumann-Morgenstern theory in situations of risk.

Let us call an act that satisfies these assumptions a lottery act. Then, on the basis of Theorem 11, we can make the following claim:

Proposition 12 *If preferences over lottery acts are linear and Archimedean then they have an expected utility representation.*

Normatively the implication is that, given a value relation on outcomes and a probability on states of the world, the only permissible actions are those that maximise the expectation of a utility measure of the value relation. Note that the utility representation is itself constrained by the assumption that preferences are linear because these imply that the manner in which outcomes are weighed against each other is sensitive in a particular way to their probabilities i.e. the assumption encodes a view about how value articulates with probability. This will be reflected, for instance, in the fact that if a preferred outcome has half the probability of a less preferred one, then its value (as measured by utility) must be twice that of the latter if the decision maker is to remain indifferent between the two.

The manner in which utility is cardinalised imposes significant constraints on how utility is interpreted. Suppose for instance that an agent is risk averse with respect to money in the sense that she prefers £50 for certain to a gamble yielding £100 with 50% probability and £0 with 50% probability. Then an expected utility representation of her preferences requires that the utility difference between receiving £50 and receiving £100 will be less than the utility difference between receiving nothing and receiving £50.

Both Arrow [3] and Sen [30] make the following objection. This way of cardinalising utility mixes up the intrinsic value to the agent of the money received with her attitude to risk taking. For it doesn't allow us to distinguish cases in which the agent prefers the £50 for certain to the gamble because of the diminishing marginal value of money from the case in which she does because she dislikes taking risks and is not willing to endanger the £50 for an even chance of doubling her money. Defendants retort that the notion of the intrinsic value being invoked in this argument lacks clear meaning. To give it a content we must be able to say how, at least in principle, we could separate the value component of preferences from the risk component that distorts it, leading to a decomposition of utility into a risk and a value component. There are several recent attempts to do so (see [11, 36] and [34]) and although it remains to be seen whether any are fully adequate, the basic conceptual point remains valid: there may be more than one type of factor contributing to an agent's preferences (apart from her beliefs).

A quite different line of criticism concerns not the interpretation of the expected utility representation, but the claims about rational preference upon which it sits. The main focus of attention in this regard has been the axiom of Independence and its violation in the so-called Allais' paradox. To illustrate the paradox, consider two pairs of lotteries yielding monetary outcomes with the probabilities given in the following table (Table 34.3).

Allais [1] hypothesised that many people, if presented with a choice between lotteries I and II would choose I, but if presented with a choice between III and IV, would choose IV. Such a pattern of choice is, on the face of it, in violation of the Independence axiom since the choice between each pair should be independent of the common consequences appearing in the third column of possible outcomes.

Table 34.3 Allais' paradox

Lottery	Probability	0.01	0.1	0.89
	I	$1000,000	$1000,000	$1000,000
	II	$0	$5000,000	$1000,000
	III	$1000,000	$1000,000	$0
	IV	$0	$5000,000	$0

Nonetheless Allais' conjecture has been confirmed in numerous choice experiments. Moreover many subjects are not inclined to revise their choices even after the conflict with the requirement of the Independence axiom is pointed out to them. So the 'refutation' seems to extend beyond the descriptive interpretation of the axiom to include its normative pretensions.

There are two lines of defense that are worth exploring. The first is to argue that the choice problem is under-described, especially with regard to the specification of the consequences. One common explanation for subjects' choices in these experiments is that they choose I over II because of the regret they would feel if they chose II and landed up with nothing (albeit quite unlikely), but IV over III because in this case the fact that it is quite likely that they will not win anything whatever they choose diminishes the force of regret. If this explanation is correct then we should modify the representation of the choice problem faced by agents so that it incorporates regret as one possible outcome of making a choice. The same would hold for any other explanation of the observed pattern of preferences that refers to additional non-monetary outcomes of choices.

The second line of defensive argument points to the gap between preference and choice. As we noted before, the specification of the choice set can influence the agent's attitudes. This is just such a case. In general the attitude we take to having or receiving a certain amount of money depends on our expectations. If we expect $100, for instance, then $10 is a disappointment. Now the expectation created by presenting the agent with two lotteries to choose from is quite different in the case where the choice is between lotteries I and II and the one in which the choice is between lotteries III and IV. In the first case they are being placed in a situation in which they can expect to gain a considerable amount of wealth, while in the second they are not. In the first they can think of themselves as being given $1000,000 and then having the opportunity to exchange it for lottery II. In the second case they can think of themselves as being handed some much lesser amount (say, whatever they would pay for lottery III) and then being given the opportunity to exchange it for lottery IV. Seen this way it is clear why landing up with nothing is far worse in the first case than in the second. It is because of what one has given up for it. In the first case landing up with nothing as a result of choosing II is equivalent to losing $1000,000 relative to one's expectations, whereas in the second case it is equivalent to losing some much smaller amount.

Both of these defences are unattractive from the point of view of constructing a testable descriptive theory of decision making under uncertainty. The first approach makes it very hard to tell what choice situation the agents face, since the description of the outcomes of the options may contain subjective elements. The second approach makes it difficult to use choices in one situation as a guide to those that will be made in another, since all preferences are in principle choice-set relative. But from a normative point of view they go some way to defending the claim that the Independence axiom is a genuine requirement of rationality.[7]

If we accept the normative validity of the Independence axiom, then we can draw the following conclusion. When the choices that we face can be represented by lotteries over a set of outcomes then rationality requires that we choose the options with maximum expected utility relative to the given probabilities of their outcomes and a given value/preference relation. What this leaves unanswered is why we should think that decision making under uncertainty can be so represented. To answer it we must return to Savage.

34.6.2 Savage' Theory

Savage [29] proves the existence of an expected utility representation of preference in two steps. First he postulates a set of axioms that are sufficient to establish the existence of a unique probability representation of the agent's beliefs. He then shows that probabilities can be used to construct a utility measure on consequences such that preferences amongst gambles cohere with their expected utilities, first on the assumption that the set of consequences is finite and then for the more general case of infinite consequences. Since the second step is essentially an application of Von Neumann and Morgenstern's theory, we will focus on the first and in particular on his derivation of a qualitative probability relation over events.

Savage takes the preference relation to be defined over a very rich set of acts, namely all functions from states to consequences. Because of its importance, I have 'promoted' the definition of the domain of the preference relation to being an additional postulate.

P0 (*Rectangular field assumption*[8]): $\Gamma = \Omega^S$
P1 (*Ordering*) $\succeq$ is (a) complete and (b) transitive.

For any events $F, G \in \mathcal{F}$, let acts $\bar{\alpha}$ and $\bar{\beta}$ be the corresponding constant acts such that for all states s, $\bar{\alpha}(s) = F$ and $\bar{\beta}(s) = G$. Given this definition it is straightforward to induce preferences over consequences from the preferences over acts by requiring that $F \succeq G$ iff $\bar{\alpha} \succeq \bar{\beta}$.

Savage's next step is to assume that the preference relation is separable across events i.e. that the desirability of a consequence of an act in one state of the world is independent of its consequences in other states. He does so by means of his famous Sure-thing principle. Consider the actions displayed in the table below.

[7] For arguments that it is not a requirement of rationality see [11] and [33].

[8] I take this term from Broome [10].

	Events	
Actions	E	E'
α	X	Y
β	X^*	Y

Then action α should be preferred to action β iff consequence X is preferred to consequence X^*. This is because α and β have the same consequence whenever E is not the case, and so should be evaluated solely in terms of their consequences when E is the case. Consequently any other actions α' and β' having the same consequence as α and β respectively whenever E is the case, and identical consequences when E is not, should be ranked in the same order as α and β. More formally:

P2 (*Sure-thing Principle*) Suppose that actions α, β, α' and β' are such that for all states $s \in E$, $\alpha(s) = \alpha'(s)$ and $\beta(s) = \beta'(s)$ while for all states $s \notin E$, $\alpha(s) = \beta(s)$ and $\alpha'(s) = \beta'(s)$. Then $\alpha \succeq \beta$ iff $\alpha' \succeq \beta'$

In view of P2 we can coherently define the conditional preference relation 'is not preferred to, given B', denoted $\succeq_B$, by $\alpha \succeq_B \beta$ iff $\alpha' \succeq \beta'$, where the acts α' and β' are as defined in P2. Given P2, it follows from this definition that the conditional preference relation is complete and transitive. This puts us into territory familiar from the discussion of conjoint additive structures. Given P0–P2, Theorem 7 implies that there exists an additive utility representation of preferences over acts that is unique up to positive affine transformation, i.e. such that the value of each act is the sum of the state-dependent utilities of its consequences.

This representation does not disentangle the contributions of the probabilities of states from the desirabilities of the consequences. To go further, assumptions that ensure the comparability of the state-dependent utilities are needed. Let us call an event $E \in \mathcal{F}$ a null event iff $\alpha \approx_E \beta$, for all $\alpha, \beta \in \Gamma$. Then Savage postulates:

P3 (*State Independence*) Let $B \in \mathcal{F}$ be non-null. Then if $\alpha(s) = F$ and $\alpha'(s) = G$ for every $s \in B$, then $\alpha \succeq_B \alpha' \Leftrightarrow F \succeq G$

The state independence assumption ensures the *ordinal* uniformity of preferences across states, but is not strong enough to ensure the *cardinal* comparability of the state-dependent utilities. The next step is the crucial one for ensuring this as well as for obtaining a probability representation of the agent's attitudes to events. First Savage defines a 'more probable than' relation, $\trianglerighteq$, on the set of events. Consider the following pair of actions:

	Events	
Action	E	E'
α	X	Y

	Events	
Action	F	F'
β	X	Y

Actions α and β have the same two possible consequences, but α has the preferred consequence whenever E is the case and β has it whenever F is the case. Now suppose that consequence X is preferred to consequence Y. Then α should be preferred to β iff E is more probable than F because the action which yields the better consequence with the higher probability should be preferred to one which yields it with lower probability. More formally:

Definition 13 (Qualitative probability) Suppose $E, F \in \mathcal{F}$. Then $E \trianglerighteq F$ iff $\alpha \succeq \beta$ for all actions α and β and consequences X and Y such that:

(i) $\alpha(s) = X$ for all $s \in E, \alpha(s) = Y$ for all $s \notin E$,
(ii) $\beta(s) = X$ for all $s \in F, \beta(s) = Y$ for all $s \notin F$,
(iii) $X \succeq Y$

In effect the circumstances postulated by this definition provides a 'test' for when one event is more probable than another. A further postulate is required to ensure that this test can be used to compare any two events in terms of their relative probability.

P4 (*Probability Principle*) $\trianglerighteq$ is complete

To apply Theorem 7, our earlier representation theorem for additive conjoint structures, we need to confirm that the derived 'more probable than' relation is not only complete, but transitive and quasi-additive. In fact this follows straightforwardly from P0 to P4 and the definition of the 'more probable than' relation. Two further structural axioms are required to ensure that the qualitative probability relation defined by P4 can be represented numerically.

P5 (*Non-Triviality*) There exists actions α and β such that $\alpha \succ \beta$.
P6 (*Non-Atomicity*) Suppose $\alpha \succ \beta$. Then for all $X \in \mathcal{F}$, there is a finite partition of S such that for all $s \in S$:

(i) $(\alpha'(s) = X$ for all $s \in A, \alpha'(s) = \alpha(s)$ for all $s \notin A)$ implies $\alpha' \succ \beta$.
(ii) $(\beta'(s) = X$ for all $s \in B, \beta'(s) = \beta(s)$ for all $s \notin B)$ implies $\alpha \succ \beta'$.

P6 is quite powerful and implies that there are no consequences which are so good or bad, that they swamp the improbability of any given event A. Nonetheless neither it nor P5 raises any pressing philosophical issues. And using them Savage proves:

Theorem 14 *There exists a unique probability function P on $\mathcal{F}$ such that for all $E, F \in \mathcal{F}$:*

$$P(E) \geq P(F) \Leftrightarrow E \trianglerighteq F$$

It is not difficult to see how in principle this theorem can serve as the basis for deriving an expected utility representation. In essence what needs to be established is a correspondence between each act α and a lottery which yields each possible consequence $C \in \Omega$ with probability, $P(\alpha^{-1}(C))$. Then since Savage's postulates for preferences over acts with a finite number of consequences imply that the induced preferences over the corresponding lotteries satisfy the Von Neumann and Morgenstern axioms, the utility of each such act can be equated with that of the expected utility of the corresponding lottery. The proof of this is far from trivial and we won't examine it here – see Savage [29] or Kreps [24] for details.[9]

34.6.3 The Status of Savage's Axioms

34.6.3.1 The Sure-Thing Principle

The most controversial of Savage's axioms is undoubtedly the Sure-thing Principle, Savage's version of the separability condition that appears with different names in different contexts. Although the Independence axiom of Von Neumann and Morgenstern's decision theory is not implied by the Sure-thing principle alone (P3 in particular is also required), the criticism based on Allais' paradox is clearly applicable here as well as are the lines of defence previously sketched. We will not repeat this discussion. But it is worth drawing attention to one further issue. As is evident from the informal presentation of the Sure-thing principle, it is essentially a principle of dominance. That is to say that its intuitive appeal rests on the thought that since only the consequences of an action matter to its evaluation, if the consequences of one act are as least as good as those of another, and are better in at least one event, then this act is better overall. But this application of Consequentialism is mistaken. For it matters not just what consequences an action has, but how probable these outcomes are and in particular how probable it makes them. Two actions can have identical consequences but if one of them brings about the better consequences with a higher probability than the other then it should be preferred.

The upshot of this is that the Sure-thing is not unconditionally valid as a principle of rationality. It is binding only if the states of the world are probabilistically independent of the acts being compared by reference to these states. But this presents Savage with a very significant problem. Amongst other things, his representation theorem is supposed to establish conditions under which a probability measure of belief can be attributed to the decision maker. But it now seems that the attribution

[9] Savage in fact introduces one further postulate necessary for the extension of the expected utility representation to infinite consequences sets. This final postulate is very much in the spirit of the Sure-thing principle and as it does not raise any additional conceptual issues, I will not state it here.

process depends on being able to establish what the decision maker's beliefs are in order to determine whether the Sure-thing principle is applicable. So Savage needs to assume precisely that which he hopes to deduce. Remarkably this fundamental difficulty has been all but ignored in the wide ranging decision-theoretic literature on belief identification.

State-Independence

The axiom of State-Independence requires that if constant act α is preferred to constant act β, given some non-null event E, then α is preferred to β, for any other non-null event F. It is not hard to produce counterexamples. Consider an act which has the constant consequence that I receive £100 and suppose I prefer it to an act with the constant consequence that I receive a case of wine. Would I prefer receiving the £100 to the case of wine given any event? Surely not: in the event of high inflation for instance, I would prefer the case of wine. One could retort that receiving £100 is not a genuine consequence since its description fails to specify features relevant to its evaluation. Perhaps 'receiving £100 when inflation is low' might be closer to the mark. But then the rectangular field assumption forces us to countenance actions which have this consequence, namely my receiving £100 when inflation is low, in states of the world in which inflation is high. Such acts seems nonsensical however and it is hard to see how anyone could express a reasonable preference regarding them.

An objection of this kind was famously made by Robert Aumann in a letter to Savage in 1971.[10] Savage's reply is interesting. He suggests that "a consequence is in the last analysis an experience" [12, p. 79], the implication being that experiences screen out the features of the world causing them and hence have state-independent desirabilities. This is unpersuasive. Even the desirability of experiences are contingent on the state of the world. On the whole I prefer that I be amused than saddened (or experience amusement to experiencing sadness), but I surely do not prefer it, given that a close friend has died. A second objection is more fundamental. To identify consequences with subjective experiences is to risk confusing the outcome of an action with one's evaluation of it. When I want to make a decision, say about whether to go for a swim, I need to know first what the outcome of this decision will be in the various possible states of the world. Then I try and evaluate these outcomes.

To the objection that his theory countenances nonsensical or impossible acts, Savage retorts that it is not necessary that the such acts "…serve something like construction lines in geometry" [12, p. 79], and that they need not be available in order for one to say whether they would be attractive or not. But he seems to under-appreciate the problem. Consider the decision whether or not to buy a life insurance policy that pays out some sum of money in the event of one's death. Now the pay-out

[10] Printed, along with Savage's letter in reply, in Drèze [12, pp. 76–81].

is not a state-independent consequence in Savage's sense, for I am not indifferent between being paid while alive and being paid while dead. However the natural refinement gives us the consequence of 'pay-out and dead' which patently cannot be achieved in any state of the world in which I am alive.

State-Dependent Utility

Although the assumption of state-independence is essential to Savage's representation theorem (and many others, including the widely used Anscombe-Aumann theory [2]), it is not intrinsic to the principle that rationality requires picking the option whose exercise has greatest expected benefit. Indeed Savage's theory can be generalised to a state-dependent version in the following way. For each state of the world $s_j$let v_j be a real-valued (utility) function on consequences measuring their desirability in that state of the world. Then the probability-utility matrix induced by the decision problem takes the form:

	States of the world		
Options	$P(\{s_1\})$	...	$P(\{s_n\})$
A^1	$v_1(C_1^1)$	...	$v_n(C_n^1)$
...	...	...	...
A^m	$v_1(C_1^m)$	...	$v_n(C_n^m)$

The expected benefit on any option is given, as before, by the expected value of the random variable which specifies its consequences. In this case this is defined by:

$$EU(A^i) = \sum_{j=1}^{n} v_j(C_j^i).P(s_j)$$

Proving the existence of such a representation is straightforward: as we observed earlier, given P0, P1 and P2 we can apply Theorem 7 to establish the existence of u_j such that $U(A^i) = \sum_{j=1}^{n} u_j(C_j^i)$ and then for any probability mass function P on the states of the worlds define $v_j := \frac{u_j}{P(s_j).}$. The problem is that the choice of probability function P here is arbitrary and there is no reason to think that it measures the decision maker's degrees of belief.[11]

[11] See, for instance, Drèze [12], Karni et al. [21] and Karni and Mongin [20] for a discussion of this issue.

34.7 Decision Theory: Evidential and Causal

This discussion of Savage's axioms reveals that three conditions must be satisfied for the maximisation of expected utility to be a rational basis for choice (assuming the absence of option uncertainty). Firstly, there must be no option uncertainty. Secondly, the desirability of each of the consequences must be independent both of the state of the world in which it obtains and the action that brings it about. And thirdly, the states of world must be probabilistically independent of the choice of action. Taken separately it is often possible to ensure that for all practical purposes these conditions are met by taking care about how the decision problem is framed. But ensuring that all three are satisfied at the same time is very difficult indeed since the demands they impose on the description of the decision problem pull in different directions. For instance option uncertainty can be tamed by coarsening the description of outcomes, but eliminating state-dependence requires refining them.

This problem provides strong grounds for turning our attention to a rival version of subjective expected utility theory that is due to Richard Jeffrey and Ethan Bolker. Jeffrey [18] makes two modifications to the Savage framework. First, instead of distinguishing between the objects of preference (actions), those of belief (events) and those of desire (consequences), Jeffrey takes the contents of all of the decision maker's attitudes to be propositions. And secondly, he restricts the set of actions to those propositions that the agent believes he can make true at will.

The first of these modifications we have already in effect endorsed by arguing that the difference between states and consequences is pragmatic rather than logical. Furthermore, if the contents of propositions are given by the set of worlds in which it is true, then Jeffrey's set of propositions will simply be Savage's set $\mathcal{F}$ of events, the only difference between the two being that there is no restriction of consequences to maximally specific propositions in Jeffrey's framework. This small modification has a very important implication however. Since states/events and consequences are logically interrelated in virtue of being the same kind of object, consequences are necessarily state-dependent. This means that Jeffrey's theory is not subject to the second of the restrictions required for Savage's theory.

The second modification that Jeffrey makes is more contentious and requires a bit of explanation. If he followed Savage in defining actions as arbitrary functions from partitions of events to consequences, the enrichment of the set of consequences would lead to an explosion in the size of the set of actions. But Jeffrey argues that many of the actions so defined would be inconsistent with the causal beliefs of the decision maker. Someone may think they have the option (which we previously named 'taking the car') of making it true that if the traffic is light they arrive on time, and if it's heavy they arrive late, but not believe that they have the option of making it true that if the traffic is light they arrive late, and if it's heavy they arrive on time. Yet Savage's rectangular field assumption requires that such options exist and that the agent takes an attitude to them. But if the agent doesn't believe that such options are causally possible, then any attitudes we elicit with regard to them may be purely artifactual.

We can look at this issue in a slightly different way. As we noted in the discussion of option uncertainty, an agent may be uncertain as to what consequences its performance yields in each state of the world. So they may not know what actions qua mappings from states to consequences are available to them. Jeffrey's solution is to conceive of an action, not as a mapping from states to consequences, but as a subjective probability distribution over consequences that measures how probable each consequence would be if the action were performed. This means that when evaluating the act of taking an umbrella for instance, instead of trying to enumerate the features of the state of the world that will ensure that I stay dry if I take an umbrella, I simply assess the probability that I will stay dry if I take the umbrella and the probability that I will get wet anyhow (even if I take it). I should then perform the act which has the greatest conditional expected utility given its performance

Two features of this treatment are noteworthy. Firstly, it is no longer required that the states of the world be probabilistically independent of the available actions. On the contrary, actions matter because they shape probabilities. This dispenses with the third constraint on the applicability of Savage's theory. Secondly, agents are not able to choose between arbitrary probability distributions over consequences but are restricted to those probability distributions that they consider themselves able to induce through their actions. To put it somewhat differently, we may think of both Savage's and Jeffrey's actions as inducing Von Neumann and Morgenstern lotteries over consequences. But Jeffrey only countenances preferences over lotteries which conform with the agent's beliefs. This solves the problem of option uncertainty by endogenising it. The agent is not option uncertain about an action because what action it is (what probability distribution it induces) is defined subjectively i.e. in terms of the agent's beliefs.

34.7.1 Desirability Representations

Let us now turn to the representation of preferences in the Bolker-Jeffrey theory. Recall that for Jeffrey the content of both beliefs and desires are propositions. To emphasise the contrast with Savage, let us model propositions as sets of possible worlds or states of the world. Then an agent's beliefs are measured, as in Savage, by a probability measure P on $\mathcal{F}$, the set of all propositions, while her degrees of desire are represented by a real valued (desirability) function V on $\mathcal{F} - \{\bot\}$, the set of non-contradictory propositions, that satisfies:

Axiom 15 (Desirability) *If* $X \cap Y = \varnothing$, *and* $P(X \cup Y) \neq 0$, *then:*

$$V(X \cup Y) = \frac{V(X).P(X) + V(Y).P(Y)}{P(X \cup Y)}$$

The notion of a desirability function on the set of propositions extends the quantitative representation of the agent's evaluative attitudes from just the maximally specific ones (that play the role of consequences in Savage's theory) to the full set of them. The basic intuition behind the extension encoded in the desirability axiom is the following. How desirable some coarse grained proposition is depends both on the various ways it could be realised and on the relative probability of each such realisation, given the truth of the proposition. For instance, how desirable a trip to beach is depends on how desirable the beach is in sunny weather and how desirable it is in rainy weather, as well as how likely it is to rain or to be sunny, given the trip.

What properties must preferences on prospects satisfy if preferences are to have a desirability representation i.e. such that $X \succeq Y \Leftrightarrow V(X) \succeq V(Y)$? There are two:

Axiom 16 (Averaging) *If $X \cap Y = \varnothing$, then:*

$$X \succeq Y \Leftrightarrow X \succeq X \cup Y \succeq Y$$

Axiom 17 (Impartiality) *If $X \cap Z = Y \cap Z = \varnothing$, $X \approx Y \not\approx Z$ and $X \cup Z \approx Y \cup Z$, then for all $Z' \in \Omega$ such that $X \cap Z' = Y \cap Z' = \varnothing$, it is the case that $X \cup Z' \approx Y \cup Z'$.*

The Averaging axiom say that if X is preferred to Y then X should be preferred to the prospect that either X or Y is the case, since the latter is consistent with Y being the case while the former is not. It has a somewhat similar motivation to the Sure-thing principle, but is much weaker. In particular it is not directly vulnerable to the Allais' paradox.

The impartiality axiom allows for a partial separation of beliefs and desires. The idea is as follows. Suppose propositions X and Y are equally preferred and that Z is some proposition disjoint from and preferred to both. Then the disjunction of X and Z will be equally preferred to the disjunction of Y and Z iff X and Y are equally probable. If X were more probable than Y then the probability of Z conditional on $X \cup Z$ would be less than the probability of Z conditional on $X \cup Y$. And so the prospect of $X \cup Z$ would be less desirable that than of $Y \cup Z$ since it would yield the more desirable prospect (Z) with lower probability.

Theorem 18 ((Bolker [6])) *Let $\mathcal{F}$ be an atomless Boolean algebra of propositions. Let $\succeq$ be a continuous weak preference order on $\mathcal{F}$. Then there exists a probability measure P and signed measure U on Ω, such that for all $X, Y \in \mathcal{F} - \{F\}$, such that $P(X) \neq 0 \neq P(Y)$:*

$$X \geq Y \Leftrightarrow \frac{U(X)}{P(X)} \geq \frac{U(Y)}{P(Y)}$$

Furthermore $P^{'}$ and U' are another such pair of measures on Ω iff there exists real numbers a, b, c and d such that (i) $ad - bc > 0$, (ii) $cU(T) + d = 1$, (iii) $cU + dP > 0$, and:

$$P^{'} = cU + dP$$

$$U' = aU + bP$$

It follows that the desirability function V defined by for all $X \in \mathcal{F}$ such that $P(X) \neq 0$, $V(X) = \frac{U(X)}{P(X)}$, represents the preference relation $\succeq$ but only up to fractional linear transformation. The uniqueness properties here are considerably weaker than in Savage's framework and this is perhaps the reason for the unpopularity of Jeffrey's theory amongst economists and applied decision theorists. It is not an insurmountable problem however for there are ways of strengthening Bolker's representation theorem, either by postulating direct probability comparisons (see Joyce [19]) or by enriching the set of propositions by conditionals (see Bradley [9] and [7]).

34.7.2 Causal Decision Theory

Jeffrey's decision theory recommends choosing the action with maximum desirability. Two closely related questions arise. Firstly, is this the same recommendation as given by Savage's theory? And secondly, if not, which is correct? The answer to the first question is less clear cut than might be hoped. On the face of it the prescriptions are different: Jeffrey requires maximisation of the conditional expectation of utility, given the performance of the action, while Savage requires maximisation of unconditional expectation of utility. But since they represent actions differently these two prescriptions are not directly in conflict. In fact, there is a way of making them perfectly compatible. The trick is to represent a Savage-type action within the Jeffrey framework by an indicative conditional of the form 'If the state of the world is s_1, then the consequence is C_1; if the state of the world is s_2, then the consequence is C_2; ...'. Then, given some reasonable assumptions about the logic of conditionals and rational preferences for their truth, the desirability of action-conditionals will just be the expected desirability of the consequence to which it refers, relative to the probability of the states with which the consequences are associated. (See Bradley [7] for details.)

On this interpretation, Savage and Jeffrey's theories are both special cases of a larger Bayesian decision theory. Most causal decision theorist reject this view and regard Savage's theory as a precursor to modern causal decision theory, which prescribes not maximisation of desirability but maximisation of causal expected utility. The distinction is brought out rather dramatically by the famous Newcomb's paradox, but since this example raises issues tangential to the main one, let us use a more banal example. Suppose that I am deliberating as to whether to eat out at Chez Posh next week. Chez Posh is very expensive, so not surprisingly the probability of being rich given that one eats there is high. I now apply Jeffrey's theory as follows. There are three prospects of interest: A: I have a good meal, B: I will be rich and C: I eat at Chez Posh. Then assuming that eating at Chez Posh guarantees a good meal:

$$V(C) = V(A \cap B \cap C).P(B|C) + V(A \cap \neg B \cap C).P(\neg B|C)$$

Since both $P(B|C)$ and $V(A \cap B \cap C)$ are high, Jeffrey's theory recommends going. But if I cannot afford to go, this will be very bad advice!

The problem is easy to spot. Although the probability of being rich given that one eats at Chez Posh is high, deciding to eat there won't make one rich. On the contrary it will considerably aggravate one's penury. So desirability is a poor guide to choice-worthiness in this case and indeed in any case when the performance of an action is evidence for a good (or bad) consequence but not a cause of it. Causal decision theory proposes therefore that actions be evaluated, not in terms of desirability, but in terms of the efficacy of actions in bringing about desired consequences.

More formally for each option A^i let P be a probability mass function on states of the world with p^i_j being the probability that s_j would be the state of the world were option A^i exercised. Let u^i_j be the utility of the consequence that results from the exercise of option A^i in state s_j. Then a decision problem can be represented by the following probability-utility matrix.

	States of the world		
Options	s_1	...	s_n
A^1	(p^1_1, u^1_1)	...	(p^1_n, u^1_n)
...	...	...	...
A^m	(p^m_1, u^m_1)	...	(p^m_n, u^m_n)

In this general case, the requirement of rationality to pick the option whose exercise has greatest expected benefit is made precise by causal decision theory, as the requirement to choose the option with maximum causal expected utility (CU), where this is defined as follows:

$$CU(A^i) = \sum_{j=1}^{n} u^i_j . p^i_j$$

In the special case when p^i_j equals $P(\{s_j\}|A^i)$ then the value of an option is given by its conditional expectation of utility, V, where this is defined by:

$$V(A^i) = \sum_{j=1}^{n} u^i_j . P(\{s_j\}|A^i)$$

This is just Jeffrey's desirability measure. So on this interpretation Jeffrey's theory is special case of causal decision theory, applicable in cases where the probability of a consequence on the supposition that an action is performed is just the conditional probability of the consequence given the action.

34.8 Ignorance and Ambiguity

The agents modelled in the decision theories described in the previous two sections are not only rational, but logically omniscient and maximally opinionated. Rational in that their attitudes – beliefs, desires and preferences – are consistent both in themselves and with respect to one another; logically omniscient because they believe all logical truths and endorse all the logical consequences of their attitudes; and opinionated because they have determinate belief, desire and preference attitudes to all prospects under consideration either because they possess full information or because they are willing and able to reach judgements on every possible contingency.

Relaxations of all of these assumptions have been studied both within empirical and normative decision theory. Firstly there is a growing literature on bounded rationality which looks at the decision making of agents who follow procedural rules or heuristics. Most of this work has descriptive intent, but some of it retains a normative element in that it seeks to show how bounded agents, with limited computational resources, should make decisions given their limitations.[12] Secondly, the problem of logical omniscience has been tackled in different ways by either modelling agents' reasoning syntactically and restricting their ability to perform inferences or by introducing possible states of the world that are subjectively possible, but objectively not.[13] Finally, there has been a long standing debate about how agents should make decisions when they lack the information necessary to arrive at precise probabilistic judgements, which we look at below. More recently this has been supplemented by a growing literature on the requirements on rationality in the absence of the completeness assumption. With some notable exceptions this literature is almost entirely focused on incomplete probabilistic information.

34.8.1 Decisions Under Ignorance

Let us consider the extreme case first when the decision maker knows what decision problem she faces but holds no information at all regarding the relative likelihood of the states of the world: a situation termed ignorance in the literature. There are four historically salient proposals as to how to make decisions under these circumstances which we can illustrate with reference to our earlier simple example of the decision as to whether to take a bus or walk to the appointment. Recall that the decision problem was represented by the following matrix.

[12] See for instance, Simon [31], Gigerenzer and Selten [15, 37] and Rubinstein [28].

[13] See for instance, Halpern [16] and Lipman [26].

	Heavy traffic	*Light traffic*
Take a bus	−2	1
Walk	−1	−1

1. *Maximin*: This rule recommends picking the option with the best worst outcome. For instance, taking the bus has a worst outcome of -2, while walking has a worst outcome of -1. So the rule recommends walking.
2. *MaxMean*: This rule recommends picking the option with the greatest average or mean utility. For instance, taking the bus has a mean utility of -0.5, while walking has a mean utility of -1. So on this rule taking the bus is better.
3. *Hurwicz Criterion*: Let Max_i and Min_i respectively be the maximum and minimum utilities of the possible outcomes of action α_i. The Hurwicz criterion recommends choosing the option which maximises the value of $h.Max_i + (1 - h)Min_i$ where $0 \leq h \leq 1$ is a constant that measures the decision maker's optimism. In our example, for instance, the rule recommends taking the bus for any values of h such that $h > \frac{1}{3}$: roughly as long as you are not too pessimistic.
4. *Minimax Regret*: Let the regret in a state of the world associated with an action α be the difference between the maximum utility that could be obtained in that state of the world, given the available actions, and the utility of the consequence of exercising α. The minimax regret rule recommends picking the action with the lowest maximum regret. For instance, in our example the regret associated with taking a bus is 1 if the traffic is heavy and 0 if its light, while that associated with walking is 0 if the traffic is heavy but 2 if it is light. So the rule recommends taking the bus.

Each of these criteria faces serious objections. Minimax Regret violates the aforementioned Independence of Irrelevant Alternatives condition and for that reason is widely regarded as normatively unacceptable (but note that we criticised this condition on the grounds that the composition of the choice set can be relevant). With the exception of Maximin, none of the rules give recommendations that are invariant under all positive monotone transformations of utilities. But in the absence of probabilistic information how are the utilities to be cardinalised? The Maximin rule seems unduly pessimistic. For instance, even if taking a bus has utility of 1000 in case of light traffic it recommends walking. The Hurwicz criterion seems more reasonable in this regard. But both it and Maximin face the objection that refinements of the decision problem produce no reassessment in situations in which it should. Consider for instance the following modified version of our decision matrix in which we have added both a new possible state of the world – medium traffic – and a new option.

	Heavy traffic	*Medium traffic*	*Light traffic*
Take a bus	−2	−2	1
Walk	−1	−1	−1
Car	−2	1	1

On both the Hurwicz criterion and the Maximin rule, taking a bus and driving a car are equally good, even though taking a car weakly dominates taking the bus. So they seem to give the wrong prescription. It is possible to modify these two rules so that they deal with this objection. For instance, the Leximin rule adds to Maximin the condition that if two options have equally bad worst outcomes then they should be compared on the basis of their second worst outcomes, and if these are equal on the basis of their third worst, and so on. A lexicographic version of the Hurwicz criterion is also conceivable. But the possibility of ties amongst worst and best outcomes pushes us in the direction of considering all outcomes. In which case we need to consider what weights to attach to them. The answer implicitly assumed by the Maxmean rule is that we should give equal weights in the absence of any information by which they can be distinguished (this is known as the Principle of Indifference). Unfortunately this procedure delivers different recommendations under different partitionings of the event space, so Maxmean too is not invariant in the face of refinements of the decision problem.

The fact that all these proposals face serious objections suggests that we are asking too much from a theory of decision making under ignorance. In such circumstances it is quite plausible that many choices will be permissible and indeed that rationality does not completely determine even a weak ordering of options in every decision problem. If this is right we should look for necessary, rather than sufficient, conditions for rational choice. I have already implicitly helped myself to an obvious candidate for such a condition – Weak Dominance – in arguing against the Hurwicz criterion. Weak Dominance says that we should not choose action α when there exists another action β such that in every state of the world s, $\beta(s) \succsim \alpha(s)$ and in at least one state of the world $\bar{s}$, $\beta(\bar{s}) \succ \alpha(\bar{s})$. But however reasonable Weak Dominance may look at first sight, it is only valid as a principle in circumstances in which states of the world are probabilistically independent of the acts. And by assumption in conditions of complete ignorance we have no idea whether this condition is met or not. It is not that we cannot therefore use dominance reasoning, but rather that it cannot be a requirement of rationality that we do. The same applies to every kind of dominance principle. And I know of no other plausible candidates for necessary conditions on rational evaluation of options other than transitivity. Since no consistent set of beliefs is ruled out in conditions of complete ignorance, it is possible that there are no further constraints on preference either.

34.8.2 Decisions Under Ambiguity

We use the term ambiguity to describe cases intermediate between uncertainty and ignorance i.e. in which the decision maker holds some information relevant to the assessment of the probabilities of the various possible contingencies but not enough to determine a unique one. Cases of this kind provided early counterexamples to expected utility theory. Consider the following example due to Daniel Ellsberg [13] in which subjects must choose between lotteries that yield monetary outcomes that depend on the colour of the ball drawn from an urn. The urn is known to contain 30 red balls and 60 balls that are either black or yellow, but in unknown proportions.

When asked to choose between lotteries I and II and between III and IV, many people pick I and IV. This pair of choices violates the Sure-thing principle, which requires choices between the pairs to be independent of the prizes consequent on the draw of a yellow ball. They are also inconsistent with the way in which Savage uses the Probability Principle to elicit subjective probabilities. For it follows from his definition of the qualitative probability relation that lottery I is preferred to lottery II iff Red is more probable than Black and that IV is preferred to III iff not-Red (i.e. Black or Yellow) is more probable than not-Black (i.e. Red or Yellow). But this is inconsistent with the laws of probability. It is this latter feature that distinguishes Ellsberg's paradox from Allais'.

One plausible explanation for the observed pattern of choices is an attitude that has since been termed 'ambiguity aversion'. The thought is that when choosing between I and II, people pick the former because this gives them $100 with a known probability (namely one-third) while lottery II yields the $100 with unknown probability or, more precisely, with a probability that could lie between zero and two-thirds. Similarly when asked to choose between III and IV they choose the one that yields the $100 with a known probability, namely IV. They make these choices because they are unable to assign probabilities to Black or to Yellow and because, *ceteris paribus*, they prefer gambles in which they know what they can expect to get over gambles that are ambiguous in the sense that they don't know what they can expect from them.

So much is largely common ground amongst decision theorists. What is much less settled are the normative and explanatory implications of the paradox. There are three views one might adopt on this question. The first is that the Ellsberg paradox shows that expected utility theory is descriptively false but not normatively so and that the observed pattern of choices is simply irrational. The second view is that a preference for betting on known probabilities over unknown ones is perfectly reasonable and hence that the paradox shows that expected utility theory is both descriptively and normatively inadequate. The third is that it shows neither inadequacy because the decision problem has not been properly represented in Table 34.4. In particular if subjects care about the range of chances of winning

Table 34.4 Ellsberg's paradox

	Ball colour		
Lottery	Red	Black	Yellow
I	\$100	\$0	\$0
II	\$0	\$100	\$0
III	\$100	\$0	\$100
IV	\$0	\$100	\$100

yielded by their choice then this property of outcomes, and not just the final winnings, should be represented.[14]

The second of these views has inspired a number of different proposals for decision rules rivalling that of maximisation of expected utility, but this literature is growing so fast that any survey is likely to find itself out of date very quickly and so I will confine myself to mentioning a few of the more salient ones. Most proposals start from the observation that the information subjects hold in Ellsberg's problem constrains them to a family of admissible probability functions on states, each assigning $\frac{1}{3}$ to Red, some value p in the interval $[0, \frac{2}{3}]$ to Black and a corresponding value $1 - p$ to Yellow. This family of probabilities, together with a utility function on the monetary prizes, then induces a corresponding family of admissible expected utilities for the alternatives under consideration. What distinguishes the various proposals is how they see subjects as using this set of expected utilities as a basis for choice.

Perhaps the predominant proposal is the Maximin EU rule (or MEU), which requires choice of the alternative with the greatest minimum expected utility. For instance, while lottery I has expected utility $\frac{1}{3} \times U(\$100) + \frac{2}{3} \times U(\$0)$, lottery II has expected utility in the range $[U(\$0), \frac{2}{3} \times U(\$100) + \frac{1}{3} \times U(\$0)]$. The minimum value for the latter is $U(\$0)$ (assuming that utility is a positive function of money), so lottery I is better according to the MEU criterion. On the other hand lottery IV is better than lottery III since it has a guaranteed expected utility of $\frac{2}{3} \times U(\$100) + \frac{1}{3} \times U(\$0)$ while III has minimum expected utility of $\frac{1}{3} \times U(\$100) + \frac{2}{3} \times U(\$0)$. So MEU prescribes just the choices observed in the Ellsberg paradox. On the other hand it also prescribes indifference between lotteries I and III, since both have minimum expected utilities of $\frac{1}{3} \times U(\$100)$, whereas most people would strictly prefer lottery III.

MEU, like the Maximin rule we looked at in the previous section, seems too extreme and there are other popular rules that allow for caution in the face of uncertainty about the probabilities which are less so. For instance the α-MEU rule prescribes choice of the alternative that maximises the α-weighted average of its minimum and maximum expected utility, where $\alpha \in [0, 1]$ is interpreted as index of the agent's pessimism. A rather different proposal is the 'smooth' ambiguity model of Klibanoff, Marinacci and Mukerji [22] which values actions in accordance with

[14] See for instance [8].

a weighted average of a concave transformation of the expected utilities, where the weights are thought of as the agent's degrees of belief for the possible probability distributions over states and the concave transformation expresses their level of aversion to ambiguity. These models are more compelling than MEU and raise interesting philosophical questions about the parameters that they introduce but have yet to receive much discussion in the philosophical literature.

Ellsberg's paradox also raises an important methodological issue. Should we regard the choices the agent makes in situations of ambiguity as expressions of her preferences or as expressions of some other non-preference reason for choice? If we take the former view then we may regard her attitudes to ambiguity as a further psychological constituent of her preferences, but leave intact the standard theory of the relation between preference and choice with its implication that preferences are complete. If we take the latter view, then we can leave in place the standard view about the relation between preference, belief and desire and treat ambiguity attitudes as additional determinants of choice. The former view is the one taken by the majority of decision theorists working in the field, perhaps because of a deep commitment to revealed preference theory. But it seems philosophically more satisfactory to regard preferences themselves as potentially incomplete whenever beliefs are less than fully determinate. But which view it is best to take depends in part on what is discovered about ambiguity attitudes: how stable they are, how responsive are they to information, as so on. So it is premature to draw any strong conclusions.

References

1. Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica, 21*, 503–546.
2. Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics, 34*, 199–205.
3. Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New York: Wiley.
4. Bernouilli, D. (1954/1738). Exposition of a new theory on the measurement of risk (L. Sommer, Trans.). *Econometrica, 22*, 23–26.
5. Binmore, K. (2009). *Rational decisions*. Princeton: Princeton University Press.
6. Bolker, E. (1966). Functions resembling quotients of measures. *Transactions of the American Mathematical Society, 124*, 292–312.
7. Bradley, R. (2007). A unified Bayesian decision theory. *Theory and Decision, 63*, 233–263.
8. Bradley, R. (2016). Ellsberg's paradox and the value of chances. *Economics and Philosophy, 32*(2), 231–248.
9. * Bradley, R. (2017). *Decision theory with a human face*. Cambridge: Cambridge University Press [Presents a theory of rationality for bounded agents under conditions of severe uncertainty].
10. Broome, J. (1991). *Weighing goods*. Cambridge, MA: Basil Blackwell.
11. Buchak, L. (2013). *Risk and rationality*. Oxford: Oxford University Press.
12. Dreze, J. (1987). *Essays on economic decisions under uncertainty*. Cambridge: Cambridge University Press.

13. * Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics, 75*, 643–669 [A very influential early critical assesment of Savage's theory].
14. Evren, O., & OK, E. (2011). On the multi-utility representation of preference relations. *Journal of Mathematical Economics, 47*, 554–563.
15. Gigerenzer, G., & Selten, R. (2002).*Bounded rationality*. Cambridge: MIT Press
16. Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior, 37*, 321–339.
17. Herstein, I. N., & Milnor, J. (1953). An axiomatic approach to measurable utility. *Econometrica, 21*(2), 291–297.
18. * Jeffrey, R. C. (1983). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press [An influential formulation of decision theory in terms of propositional attitudes].
19. * Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press [The most complete formulation and defence of causal decision theory].
20. Karni, E., & Mongin, P. (2000). On the determination of subjective probability by choice. *Management Science, 46*, 233–248.
21. Karni, E., Schmeidler, D., & Vind, K. (1983). On state-dependent preferences and subjective probabilities. *Econometrica, 51*, 1021–1031.
22. Klibanoff, P., Marinacci. M., & Mukerji, S. (2005) A smooth model of decision making under ambiguity. *Econometrica, 73*, 1849–1892.
23. Krantz, D., Luce, R. D., Suppes, P., & Tversky, A. (1971). *The foundations of measurement. Volume 1. Additive and polynomial representations*. New York: Academic.
24. * Kreps, D. M. (1988). *Notes on the theory of choice*. Boulder/London: Westview Press [Excellent advanced introduction to decision theory].
25. Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy, 59*, 5–30.
26. Lipman, B. (1998). Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality. *Review of Economic Studies*, 339–361.
27. * Ramsey, F. P. (1926). Truth and probability. In D. H. Mellor (Ed.), *Philosophical papers*. Cambridge: Cambridge University Press, 2008 [Classic essay on the foundations of subjective probability and expected utility].
28. Rubinstein, A. (1998). *Modeling bounded rationality*. Cambridge: MIT Press
29. * Savage, L. J. (1954/1972). *The foundations of statistics* (2nd ed.). New York: Dover [The most influential formulation of subjective expected utility theory].
30. * Sen, A. K. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day [Classic textbook on choice functions and social choice].
31. Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.
32. Stalnaker, R. ([1972]/1981b). Letter to David Lewis. In W. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance, and time* (pp. 151–152). Dordrecht: Reidel.
33. Stefánsson, H. O., & Bradley, R. (2015). How valuable are chances? *Philosophy of Science, 82*(4), 602–625.
34. Stefánsson, H. O., & Bradley, R. (Forthcoming 2017). What is risk aversion? *British Journal of Philosophy of Science*.
35. * von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton: Princeton University Press [Classic formulation of expected utility theory].
36. Wakker, P. (2010). *Prospect theory: For risk and uncertainty*. Cambridge: Cambridge University Press.
37. Weirich, P. (2004). *Realistic decision theory: Rules for nonideal agents in nonideal circumstances*. New York: Oxford University Press